AD_____

Award Number: **W81XWH-11-1-0584**

TITLE: **Neural Basis of Empathy and Its Dysfunction in Autism Spectrum Disorders (ASD)**

PRINCIPAL INVESTIGATOR: **Michael L. Platt, PhD**

CONTRACTING ORGANIZATION: **Duke University**

REPORT DATE: **August 2012**

TYPE OF REPORT: Revised Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
August 2012	Revised Annual	01 AUG 2011 – 31 JUL 2012

4. TITLE AND SUBTITLE
Neural Basis of Empathy and its Dysfunction in Autism Spectrum Disorders (ASD)

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0584

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Steve W. C. Chang, Ph.D.

Michael Platt, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Duke University
Durham NC 27705-4677

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited (11/29/2012)

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The most human of emotions are defined by how we feel about others. Such other-regarding preferences (ORP), in the parlance of economics, both permit and necessitate the institutions that form the core of modern society. Abnormal functionalities associated with ORP in the brain are also thought to be a critical component underlying neuropsychiatric disorders marked by social deficits. The primary objectives of the research project are to develop an animal model of ORP (Objective 1), and investigate the functional contributions of the orbitofrontal cortex (Objectives 2-3) and neuropeptide oxytocin (OT) (Objective 4). We have successfully developed a rhesus macaque model of ORP and demonstrated that rhesus monkeys care about rewards delivered to another monkey (Chang et al., 2011a). We then for the first time demonstrated that inhaled OT using a pediatric nebulizer effectively reaches the central nervous system of rhesus macaques, providing a proof of concept for a clinical OT treatment method well-tolerated by children. Subsequently, we showed the inhaled OT promotes prosocial preferences and social gaze rates (Chang et al., 2012). Finally, we have recorded single neuron activity from OFC during the ORP task and found that OFC neurons are predominantly involved in computing rewarding events of self (Chang and Platt, 2011b; Chang et al., under review). Our results begin to elucidate neural mechanisms underlying ORP and open up future research aimed at understanding how therapeutic OT treatments in individuals with autism spectrum disorders work in their brains.

15. SUBJECT TERMS
social decision-making; oxytocin; vicarious reinforcement; rhesus macaques; the orbitofrontal cortex; social preference; autism spectrum disorders; prosocial behavior; other-regarding preference; social gaze

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	98	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Page

INTRODUCTION

Autism spectrum disorders (ASD) differ from other developmental disorders in that children with ASD show little interest in other people. This lack of interest is associated with other complex social deficits, including empathy and shared attention, thus further disrupting the capacity to engage in normal social interactions. Research findings suggest that social problems in ASD derive, in part, from dysfunction in the neural circuits that motivate the other-regarding behaviors that shape normal social interactions. Other-regarding preferences (ORPs) describe a concern for the welfare or the benefit of others. ORPs may rely on empathy, a social-cognitive capacity severely compromised in ASD, and pathological deficits of empathy in ASD may result from a failure to understand others' internal states. Accumulating evidence implicates orbitofrontal cortex (OFC) dysfunction in the pathophysiology of ASD. Furthermore, oxytocin (OT), a neuromodulatory hormone implicated in social behavior in mammals, has also been implicated in the etiology of ASD. Understanding the neuronal properties of OFC neurons and demonstrating OT-induced changes in ORP in a rhesus macaque model will significantly advance our understanding of social processing in both healthy and ASD brains. Our research aims to develop a non-human primate model for ORP specifically designed to probe these mechanisms in healthy individuals, the neuronal mechanisms involved in the expression of ORPs, and the efficacy of pharmacological OT therapies designed to enhance social interaction in ASD.

BODY

Objectives specified in the approved Project Narratives

- Objective 1: Develop an animal model of ORPs.
- Objective 2: Determine how OFC neurons mediate ORPs.
- Objective 3: Determine the effects of OFC perturbations on ORPs.
- Objective 4: Determine whether OT can enhance positive ORP.

By tasks indicated in Statements of Work (SOW)

Task 1. Characterize neural responses in the orbitofrontal cortex (OFC)

We have completed this task. Prior to the beginning of investigating neuronal correlates of ORPs in the primate OFC, we extensively characterized the social behaviors of rhesus macaques in the ORP task (see the behavioral setup and tasks in Appendix 1). We showed that rhesus macaques (*Macaca mulatta*) spontaneously derive vicarious reinforcement from observing rewards given to another monkey, and this motivates them to subsequently deliver or withhold rewards from the other monkey depending on their social relationship and other aspects of the context. We exploited both Pavlovian and instrumental conditioning procedures to associate rewards to self (donor monkey) and/or rewards to another monkey (recipient monkey) with visual cues. Donor monkeys made more errors (i.e., incomplete trials) in the instrumental trials when cues predicted reward to recipients compared to when cues predicted reward to themselves, but made even more errors when cues predicted reward to no one. In subsequent preference tests between pairs of previously conditioned cues, the donor monkeys preferred cues paired with reward to the recipients over cues paired with reward to no one. By contrast, the donor monkeys preferred cues paired with reward to self over cues paired with reward to both monkeys delivered at the same time. Rates of looking at recipients strongly predicted the strength and valence of vicarious reinforcement, suggesting a role for attention in gating social preferences. Furthermore, the preference to donate was enhanced by greater familiarity between the two animals, and is abolished if the recipient was replaced with a juice collection bottle, demonstrating the fundamentally social nature of the task. These patterns of behavior are consistent with vicarious reinforcement derived from observing another individual receive a reward. Vicarious reinforcement in rhesus macaques may play a critical role in shaping cooperation and competition, as well as motivating observational learning and group coordination, much as it does in humans and perhaps other highly social species. We propose that vicarious reinforcement signals mediate these behaviors via neural circuits involved in reinforcement learning and decision-making that are shared by rhesus macaques and humans.

We next recorded the activity of 85 single OFC neurons in two donor monkeys performing the ORP task. **Figure 1a** illustrates the OFC region of the brain targeted with electrodes on a representative coronal slice from structural magnetic resonance scans. **Figure 1b** shows a typical OFC neuron that preferentially encoded juice rewards received by the donor monkey (i.e., *self* rewards). On choice trials, this neuron responded more strongly for *self* rewards than for the alternatives on both *Self:Neither* (i.e., choosing between rewards to self and neither) and *Self:Other* (i.e., choosing between rewards to self and other) trials. Activity associated with *self* rewards did not differ between *Self:Neither* and *Self:Other* (7.00 ± 0.47, 7.03 ± 0.46 sp/s, respectively, $P = 0.97$, Welch two sample *t*-test), but it significantly exceeded the cell's activity for *other* and *neither* rewards on *Other:Neither* trials (i.e., choosing between delivering rewards to the recipient [*other* rewards] and delivering rewards to neither [*neither* rewards]) (3.06 ± 0.40, 1.85 ± 0.42 sp/s, respectively; both $P < 0.0001$). On cued trials, this

neuron fired most to *self* rewards compared to both *other* and *neither* rewards (both $P < 0.0001$, Welch two sample *t*-test), but did not differentiate between *other* and *neither* rewards ($P = 0.25$) (**Fig. 1b**).

During reward delivery, the OFC population predominantly encoded *self* rewards compared to *other* or *neither* rewards. The response for *self* rewards was 30% greater compared to *other* rewards (0.30 ± 0.09, $P < 0.005$, paired *t*-test), and the bias for *self* rewards over *neither* rewards was greater by 17% (0.17 ± 0.08, $P < 0.05$, paired *t*-test) (**Fig. 1c**, **2a**). Population activity for *other* and *neither* rewards did not differ in OFC (0.08 ± 0.06, $P = 0.20$, paired *t*-test) (**Fig. 1c**, **2a**). On cued trials, the *self* reward bias over *other* rewards was absent (0.19 ± 0.16, $P = 0.24$, paired *t*-test), and the *self* reward bias was only weakly present over *neither* rewards (0.26 ± 0.15, $P < 0.08$). On cued trials, the population did not distinguish *other* from *neither* rewards ($P = 0.33$, paired *t*-test) (**Fig. 1c**). **Figure 3** shows population responses aligned to the times of decision-making and cue offset. These results indicate that OFC neurons predominantly compute rewards received by donor monkeys, and this information is encoded more faithfully when monkeys actively choose whom to reward. In the population, reward epoch responses differed significantly for a large number of neurons depending on reward outcome (57%), trial type (45%) and reward volume (24%) (analysis of variance, ANOVA, Methods; **Table 1**). Based on the statistical significance (ANOVA) during choice and reward epochs, we classified individual neurons as self-referenced, other-referenced, both-referenced, or unclassified. When considering the proportion of different cell types among the classified neurons based on this scheme, 78% of OFC neurons were self-referenced, whereas only 10% were other-referenced and 12% were both-referenced (both $P < 0.0001$, χ^2 test) (**Fig. 2b**).

As a complementary measure of neuronal selectivity, we also examined whether lower trial-to-trial variability was associated with preferred reward outcomes in these neurons. We tested whether the coefficient of variation in firing rate (CV; Methods) was systematically lower for *self* reward outcomes compared to *other* and *neither* reward outcomes. We indeed found this to be the case. The OFC population showed a lower CV for *self* rewards compared to others (*received* – *omitted*, -0.12 ± 0.04 [mean ± s.e.m.], $P < 0.01$, one-sample *t*-test) (**Fig. 4**). Thus, the most robust responses of neurons were also the most reliable.

In the ORP task, donors were allowed to shift their gaze at the recipient. To rule out the possibility that preferential reward responses we observed were simply driven by where the donors looked on a given trial, we compared reward epoch responses on trials *with* and *without* gaze shifts to the recipient. We found no systematic differences in the reward responses of OFC neurons at the population level ($P > 0.20$, Wilcoxon signed rank test) (**Fig. 5**). Thus, the reward-related responses in OFC cannot be simply explained by preparation to look at the recipient or elicited as a consequence of inspecting the recipient.

To test whether different neuronal frames of reference (self-, other-, and both-referenced) were anatomically segregated within OFC, we applied principal component analysis on recording coordinates to identify and test the distributions in the major axis with the largest dispersion within three-dimensional space. We did not observe any systematic anatomical clustering amongst different frames of reference in OFC, indicating that self-, other-, and both-referenced neurons within OFC were intermingled (all $P > 0.56$, Wilcoxon rank sum test). **A paper based in part upon this work is *in press* at Nature Neuroscience.**

Methods

General and behavioral procedures

All procedures were approved by the Duke University Institutional Animal Care and Use Committee, and were conducted in compliance with the Public Health Service's Guide for the Care and Use of Laboratory Animals.

Two donor (MY and MO) and five recipient monkeys (*Macaca mulatta*) participated in the experiments. Donors (both males) and recipients (four males, one female) were unrelated and not cagemates. Donors were housed in a colony with 12 other rhesus macaques, some of which were pair-housed. All the male monkeys reside in this colony, and the one female monkey resided in the adjacent colony with only females. For all monkeys, a sterile surgery was performed prior to experiments to implant a head-restraint prosthesis (Crist Instruments) using standard techniques. Donor monkeys were trained on the ORP task in the presence of a recipient. Subsequently, a second surgery was performed on donors to implant a recording chamber (Crist) providing access to OFC. All surgeries were performed under isoflourane anesthesia (1-3%), and the recording chambers were regularly cleaned, treated with antibiotics and sealed with sterile caps.

Horizontal and vertical eye positions were sampled at 1000 Hz using an infrared eye monitor camera system (SR Research Eyelink). Visual stimuli were controlled by PsychToolBox and Matlab (Mathworks). Donors and recipient monkeys both sat in primate chairs (Crist), 100 cm from one another at a 45-degree angle. Each monkey had his own computer monitor which displayed identical visual stimuli. Both the donor and recipient monkeys had their own juice tube from which juice drops were delivered. In order to prevent monkeys from forming secondary associations of solenoid valve clicks or the sound of the recipient drinking the juice reward, the solenoid valves were placed in another room and white noise was also played in the background. For example, experimenters were unable to hear solenoids anywhere inside the recording room. Critically, a separate solenoid designated for *neither* rewards was also placed outside the room; it only produced clicks but delivered no fluid. The face region of the recipient, with respect to the gaze angle of the donor (horizontal and vertical eye positions), was mapped out empirically prior to the experiments. The frequency with which donors looked at recipients was computed from counting the number of gaze shifts to the recipient's face (± 8.5˚ from the center of the face) (Chang et al., 2012a). A large window was used to capture gaze shifts that were brief in duration and large in magnitude and often directed at varying depths (e.g., eyes, mouth).

Monkeys performed the task to obtain drops of cherry- or orange-flavored juice. Donors began a trial by shifting gaze (± 2.5˚) to a central stimulus (0.5˚ x 0.5˚), and maintained fixation for 200 ms. On the majority of sessions (see above), the reward magnitude at stake (0.1 – 2.4ml) on each trial was cued by the position of a horizontal bisecting line (200ms), indicating the percentage of the maximum possible volume. There were two kinds of trials, *choice trials* and *cued trials*. Following a variable delay of 300, 500, or 700 ms, choice and cued trials were presented at equal probabilities and randomly interleaved. On choice trials, two visual targets (4˚ x 4˚) appeared at two random locations 7˚ eccentric in the opposite hemifield. Donors shifted gaze to one target (± 2.5˚) to indicate a decision within the maximum allowed time of 1.5 s relative to the stimulus onset. The pair of stimuli appearing on a given trial was drawn from the set of three stimuli, which were pseudorandomly selected. On cued trials, donors maintained fixation (± 2.5˚) while a cue (4˚ x 4˚) appeared centrally to the screen for 500 ms. Cues indicating rewards for the donor, recipient or neither monkey occurred with equal frequency and pseudorandomly determined. Reward onset was followed by a variable 0 – 900 ms delay, from the time of either making a choice (choice trials) or cue offset (cued trials). Donors were free to

look around the room during this delay and for another one second after reward delivery. Reward delivery was followed by an intertrial interval of 700, 1,000, or 1,300 ms. Upon making an error, both monkeys received visual feedback (a white rectangle, 10˚ x 10˚) followed by a 5s time out before the next trial began.

Recording procedures

All recordings were made using extracellular tungsten electrodes (FHC). Single electrodes were lowered using a hydraulic microdrive system (Kopf Instruments, or FHC). Single-unit waveforms were first isolated, and action potentials collected, using a 16-channel recording system (Plexon, Inc.). In order to guide the placement of recording tracks and localize recording sites, we acquired structural magnetic resonance images (MRI) (3T, 1 mm slices) of each donor's brain prior to the experiments (e.g., see **Fig. 1a**). Detailed localizations were made using Osirix-viewer. We also confirmed that electrodes were in OFC by listening to grey-matter and white-matter associated sounds while descending the electrodes down to OFC. OFC neurons were recorded from Brodmann areas 13m and 11 (**Fig. 1a**). A total of 85 OFC neurons (MY: 46, MO: 39) were included. These neurons were selected for recording based solely on the quality of isolation upon encountering. For a small subset of the data (27%), data were collected in a task with a fixed reward size (typically 1.0ml per successful trial). For the majority of the cells (42%), data were either collected in a task with the magnitude cue, or both with and without the magnitude cue (i.e., two or more consecutive blocks per cell) (31%). We combined the two types of data in our analyses unless otherwise specified.

Data analysis

Data from each cell consisted of firing rates during 440 ± 13 (±217) (median ± s.e.m. (±s.d.)) trials. The monkeys performed the task well, as evidenced by a high percentage of completed trials even on trials in which they did not receive juice reinforcement (e.g., also see Chang et al., 2012a).

Choice preference indices were constructed as contrast ratios (Eq. 1) (Chang et al., 2012a).

$$Preference\ Index = \frac{R_A - R_B}{R_A + R_B}. \quad (1)$$

R_A and R_B values were the frequency of making particular choices. For *Self:Other* trials, R_A and R_B were number of choices to reward *other* and *self*, respectively. For *Other:Neither* trials, R_A and R_B were number of choices to reward *other* and *neither*, respectively. Finally, for *Self:Neither* trials, R_A and R_B were number of choices to reward *neither* and *self*, respectively. Indices therefore ranged from -1 to 1, with 1 corresponding to always choosing to donate on *Other:Neither* trials and *Self:Other* trials, and always choosing not to reward self on *Self:Neither* trials. An index of -1 corresponds to the opposite, generally stated as choosing not to donate to the other monkey or choosing to reward oneself. Values of 0 indicated indifference. For constructing neuronal preferences, we simply substituted the choice frequency with neuronal firing rates associated with making specific decisions. Response times, the time from the onset of choices to movement onset, were computed using a 20°/sec velocity threshold criterion (Chang et al., 2012a).

Firing rates were computed during the reward epoch (from 50 to 600ms from reward onset) as well as for the choice epoch (from -100 to 400ms from making a choice). For the

population analyses, we normalized reward epoch firing rates to the average baseline rates for each reward outcome (300 ms interval prior to fixation onset). Using marginally different time windows and different normalization methods all resulted in similar conclusions. Coefficients of variation (CV) were calculated for each neuron based on the standard deviation (σ) and mean (μ) using the spike rates (sp/s) from the reward epoch (Eq. 2):

$$CV = \frac{\sigma}{\mu} \quad (2)$$

We used analysis of variance (ANOVA) to classify the reward response selectivity of individual neurons per individual cells. Two-factor ANOVA was used to classify the selectivity of reward outcome (*self*, *other*, or *neither*) and trial type (choice or cued). Three-factor ANOVA was used to classify the selectivity of reward volume (binned into small, medium, large) for the cells that were collected in the task with a magnitude cue. Statistical significance for each reward type was computed by Tukey HSD test. Across all analyses, using slightly different epoch durations for neuronal data analyses led to similar results and conclusions.

Classification of cell types by significant reward specificity

Based on Tukey HSD tests, we classified cells into the following categories: self-referenced, other-referenced, both-referenced (mirror), and unclassified. These categories do not imply functional roles but indicate that firing rates were significantly different based on reward outcomes. A neuron was referred as self-referenced if the responses of the neuron were significantly different ($P < 0.05$) between *self* and *other* rewards as well as between *self* and *neither* rewards, but not different between *other* and *neither* rewards. A neuron was referred as other-referenced if the responses of the neuron showed significant differences in firing rates between *self* and *other* rewards as well as between *other* and *neither* rewards, but not different between *self* and *neither* rewards. Finally, a neuron was referred as both-referenced or mirror if the responses of the neuron showed significant differences in responses between *self* and *neither* rewards as well as *other* and *neither* rewards, but not different between *self* and *other* rewards. Neurons not belonging to one of these categories were considered as unclassified. Applying slightly different criteria or differently configured ANOVA variables did not change the overall results.

Task 2. Characterize behavior after muscimol inactivation

We plan to reversibly inactivate OFC neurons with muscimol next year and examine the effects on performance on the ORP task and social attention.

Task 3. Determine behavioral response to microstimulation

We plan to microstimulate OFC neurons this year and examine the effects on performance on the ORP task and social attention.

Task 4. Examine the effect of oxytocin (OT) on task performance

We have completed this task. We showed that inhaled OT (using pediatric nebulizer) penetrates the central nervous system and subsequently enhances the sensitivity of rhesus macaques to rewards occurring to others as well as themselves in the ORP task. Roughly 2

hours after inhaling OT, donor monkeys increased the frequency of prosocial choices associated with reward to another monkey (i.e., recipient monkey) when the alternative was to reward no one. OT also increased attention to the recipient monkey (i.e., number of gazes directed at the recipient following decisions) as well as the time it took to make such a decision (i.e., eye movement choice reaction times). In contrast, within the first 2 hours following inhalation, OT enhanced selfish choices associated with delivery of reward to self over a reward to the other monkey, without influencing attention or decision reaction times. Despite the differences in species typical social behavior between rhesus macaques and more egalitarian and monogamous species (like prairie voles and humans), exogenous, inhaled OT causally promotes social donation behavior in rhesus monkeys when there is no perceived cost to self. These findings potentially implicate shared neural mechanisms and validate the use of inhaled OT as a potential therapeutic for enhancing social attention and prosocial behavior in ASD. **The detailed methods, results, and figures can be found in the Appendix 1 and were published in Proceedings of the National Academy of Sciences in 2012 (Chang et al., 2012a).**

KEY RESEARCH ACCOMPLISHMENTS

- Development of intranasal oxytocin (OT) protocol in rhesus monkeys with a confirmation that the method effectively delivers OT to the central nervous system (Meeting Presentations, Paper Published: Chang et al., 2012a)

- Demonstration of OT-induces changes in ORPs in rhesus monkeys (Presented at multiple meetings) (Meeting Presentations, Paper Published: Chang et al., 2012a)

- Recording of neuronal activity from the orbitofrontal cortex (OFC) during the ORP task (Meeting Presentations, Paper currently *in press*: Chang et al., *in press*)

- Extension of the current experiments to other prefrontal brain regions (the sulcus and gyrus of the anterior cingulate cortex) to better understand how ORP-related signals differ across different parts of the prefrontal cortex (Acquired new funding for future research [see REPORTABLE OUTCOMES])

REPORTABLE OUTCOMES

A. Publications

1. Chang SW, Gariépy JF, and Platt ML (*in press)* Neuronal reference frames for social decisions in primate frontal cortex. *Nature Neuroscience.*

2. Gariépy JF, Chang SW and Platt ML (*in press*) Brain games: Toward a neuroecology of social behavior. Invited commentary in *Beh. Brain. Sci., in press.*

3. Chang SW, Barter JW, Ebitz RB, Watson KK and Platt ML (2012a) Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques (*Macaca mulatta*). *Proc Natl Acad Sci*, 109, 959–964.

4. Chang SW, Barack DL and Platt ML (2012b) Mechanistic classification of neural circuit dysfunctions: Insights from neuroeconomics research in animals. *Biol. Psychiatry*, 72:101–106.

B. Meeting Abstracts

1. Chang SW, Gariépy JF, and Platt ML. Differential encoding of social decision outcomes by neurons in primate orbitofrontal cortex, dorsal anterior cingulate cortex and anterior cingulate gyrus. *Society for Neuroscience (New Orleans, LA), 2012 (Talk) & Contributed Talk at Society for Social Neuroscience, 2012*

2. Chang SW, Gariépy JF, and Platt ML. Neuronal reference frames for social decisions in primate prefrontal cortex. *Organization for Computational Neuroscience (Atlanta, GA), 2012 (Poster)*

3. Platt, ML. Neuronal basis of giving and receiving. *Organization for Computational Neuroscience (Atlanta, GA), 2012, invited talk.*

4. Chang SW and Platt ML. Differential coding of egocentric and allocentric reward outcomes during social interaction in primate ACC and OFC. *Society for Neuroscience (Washington, DC), 2011 (Talk)*

5. Chang SW, Barter JW, Ebitz RB, Watson KK and Platt ML. Oxytocin promotes prosocial decisions in rhesus macaques. *Society for Neuroscience (Washington, DC), 2011 (Poster)*

6. Chang SW, Barter JW, Ebitz RB, Watson KK and Platt ML. Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques *(Macaca mulatta).* Workshop on the Biology of Prosocial Behavior at Emory University *(Atlanta, GA), 2011 (Poster)*

C. Research Support (built upon this award)

1. NIH/NIMH 2/21/12 - 11/30/16
   R01 MH095894-01 (*Platt*)
   Neuronal basis of vicarious reinforcement dysfunction in autism spectrum disorder
   The goal of this project is to understand the role of prefrontal cortex in mediating vicarious reinforcement during reward allocation decisions

2. Duke Department of Neurobiology 6/01/11 – 5/31/12
   Postdoctoral Training Award in Fundamental & Translational Neuroscience
   NIH/NINDS T32 NS051156-07 (*Chang*)
   Neural basis of other-regarding preference
   The goal of this project is to understand the role of anterior cingulate cortex and orbitofrontal cortex during reward allocation decisions

3. NIH/NIMH Pending
   NIH K99/R00 Pathway to Independence (*Chang*)
   Role of oxytocin in the amygdala-prefrontal network during social decision-making
   The goal of this project is to undergo extensive training in neuroendocrinology, and study the mechanisms underlying oxytocin-mediated neural processing across amygdala and prefrontal neurons in social decision-making.

D. Graduate student mentoring

1. A successful rotation project for a Duke Cognitive Neuroscience PhD candidate, Amy A. Winecoff.
2. A successful rotation project for a Duke Cognitive Neuroscience PhD candidate, Joseph W. Barter, resulting in a second authorship in Chang et al., 2012a.
3. Mentored Jean-Francois Gariépy, visiting graduate student from U. Montreal, resulting in the development of a novel social interaction task in two rhesus monkeys and the neuronal recording of strategic interaction signals from the dorsolateral prefrontal cortex.

CONCLUSION

Other-regarding preferences (ORPs) are critical for normal social behavior, and the neural mechanisms underlying ORPs may be disrupted in neuropsychiatric disorders marked by social deficits, including the autism spectrum disorders (ASD). Both motivation-related processing in the brain and the neuropeptide oxytocin (OT) have been implicated in ASD. However, the neural mechanisms underlying ORPs remain elusive, partly due to the lack of a good animal model for studying complex social behavior. To address this gap, we developed a novel social interaction task involving two rhesus macaques, and have investigated the role of OFC neurons, previously implicated in motivation and decision-making, as well as the neuropeptide OT, which has previously been implicated in social preferences, during the expression of ORPs.

We found that rhesus monkeys care about what happens to others, as indicated by their preference to deliver juice rewards to a recipient monkey over no one and the increased social looking behavior associated with such prosocial decisions. Furthermore, prosocial preferences are enhanced following inhalation of OT by monkeys. Neuronal recording from OFC revealed a strong preference for computing rewards delivered to self compared to rewards delivered to either another monkey or no one, indicating that OFC neurons track self motivation during social interactions. Furthermore, this *self* reward preference in OFC was more faithfully encoded following active decisions by donor monkeys, consistent with a motivational function of OFC with respect to self. Our results begin to reveal how the primate brain makes decisions during social interaction with other individuals.

OT has been evaluated for potential therapeutic use in clinical conditions marked by social deficits, such as ASD, antisocial personality disorder, and schizophrenia. Notably, the intranasal nebulization method we developed, demonstrated the efficacy of, and applied here is well-tolerated by children for delivery of other therapeutics (e.g., albuterol), thus opening up avenues for early OT intervention in neuropsychiatric conditions with social deficits.

Our findings provide new opportunities for uncovering the neurophysiological and neuroendocrinological mechanisms underlying complex social behavior in a species much more closely related to humans than mice or rats. Rhesus monkeys have long served as the preferred model species for probing the neural mechanisms underlying complex cognition. Given the strong similarities in social behavior and cognition, together with remarkable homologies in neural circuitry, the rhesus macaque provides a powerful model for probing the neurobiological mechanisms of social interactions in people.

Medical Implications ("So What" section)

Our work holds promise both for understanding the basic mechanisms that support complex social behavior and translating that knowledge into improved treatment for social dysfunction in ASD. In particular, our work tests the idea that empathy derives from the activation of neural circuits that process primary emotions or feelings, such as reward or punishment, merely by observing the same things happen to other people. OT therapy for ASD and other neuropsychiatric disorders is currently being explored in clinical trials, despite uncertainty regarding the exact mechanism of action in the brain or the long-term consequences of use. By testing this drug in an animal model, we can directly confirm efficacy, efficiency, and long-term safety. Clinicians can use this information to directly inform therapeutic interventions in ASD. We have already demonstrated, for the first time in any species, that inhaled OT is

taken up by the central nervous system—an important prerequisite for exploring further clinical opportunities.

Our work promises ancillary benefits as well. Our findings regarding the functional role of the OFC will also be of use in clinical contexts. The precise way OFC contributes to social behavior remains a mystery. By delineating the neuronal properties of neurons in OFC, we can further understand how this brain area contributes to the expression of complex social behavior as well as its dysfunction. Such insights may prove invaluable in the diagnosis and treatment of social behavioral disorders that accompany head trauma, with potentially important implications for veterans of US armed forces returning from the battlefield suffering from traumatic brain injuries and attendant problems in adjusting to civilian life and society.

Ultimately, the results of these experiments will inform interventions for social disorders, on both the pharmacological and behavioral levels, and significantly improve the lives of people living with ASD and other individuals struggling with normal social life.

REFERENCES

Chang SW, Barter JW, Ebitz RB, Watson KK and Platt ML (2012a) Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques (Macaca mulatta). *Proc Natl Acad Sci*, 109, 959–964.

Chang SW, Gariépy JF, and Platt ML (*in press)* Neuronal reference frames for social decisions in primate prefrontal cortex.

SUPPORTING DATA (1 Table and 5 Figures)

Table 1

Table 1. Classification of the reward type, trial type, and reward size selectivities at the level of individual neurons from OFC based on analysis of variance.

Area	Proportion of significant neurons by factors (reward epoch)		Proportion of significant neurons between different rewards (reward epoch)		Proportion of reference frame types (reward or choice epoch)	
OFC	*Reward Outcome*	57% (*n*=85)	*Self vs. Neither*	37% (*n*=85)	*SELF frame of reference*	38% (*n*=85) [**78%**]
	Trial Type	45% (*n*=85)				
	Reward Volume	24% (*n*=62)	*Self vs. Other*	42% (*n*=85)		
	Reward Outcome x Trial Type	37% (*n*=85)			*OTHER frame of refernece*	5% (*n*=85) [**10%**]
			Other vs. Neither	14% (*n*=85)		
	Reward Volume x Reward Outcome	10% (*n*=62)				
			Self (Self:Other) vs. Self (Self:Neither)	13% (*n*=85)	*BOTH frame of reference (Mirror)*	6% (*n*=85) [**12%**]
	Reward Volume x Trial Type	10% (*n*=62)				
	Reward Outcome x Trial Type x Reward Volume	11% (*n*=62)				
					Unclassified	51% (*n*=85)

The percentages shown inside the brackets on the 4th column show the proportions out of *classfied* neurons. Significance in all panels was based on $P < 0.05$ (analysis of variance and tukey HSD tests).

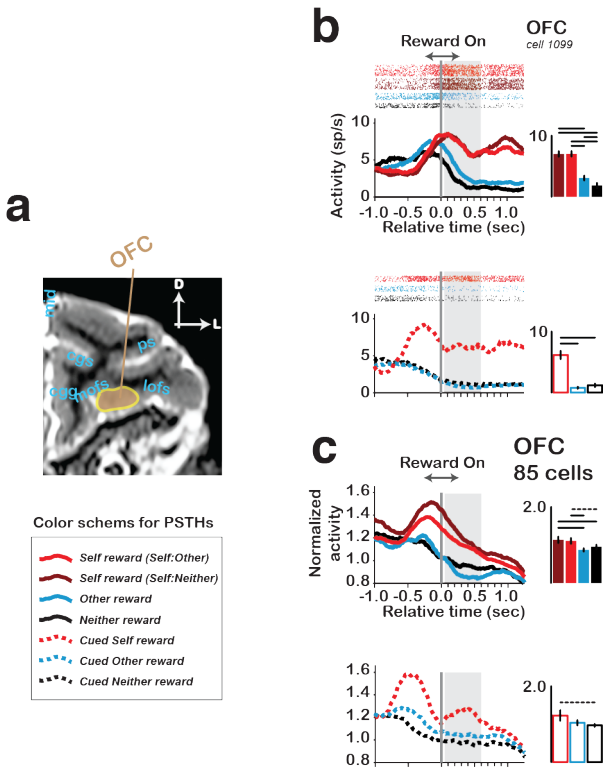


Figure 1. OFC recording sites, and single and population OFC responses during the ORP task. (**a**) Structural magnetic resonance image (MRI) from donor MO, with example electrode paths shown for OFC. (**b**) Mean responses (peri-stimulus time histograms [PSTHs]) and spike rasters for a *self* reward preferring OFC neuron. Choice trials: upper, solid traces. Cued trials: lower, dashed traces. Data are aligned to reward onset for each reward outcome. Histograms on right show mean ± s.e.m. activity from reward epoch (grey box). Color codes for PSTH traces and histograms are shown below. (**c**) Normalized reward epoch responses for 85 OFC neurons. Same format as in **b**. In all histograms, the horizontal lines above different conditions indicate significance differences (solid, $P < 0.05$ by paired *t*-test; dashed, $P < 0.05$ by bootstrap test).

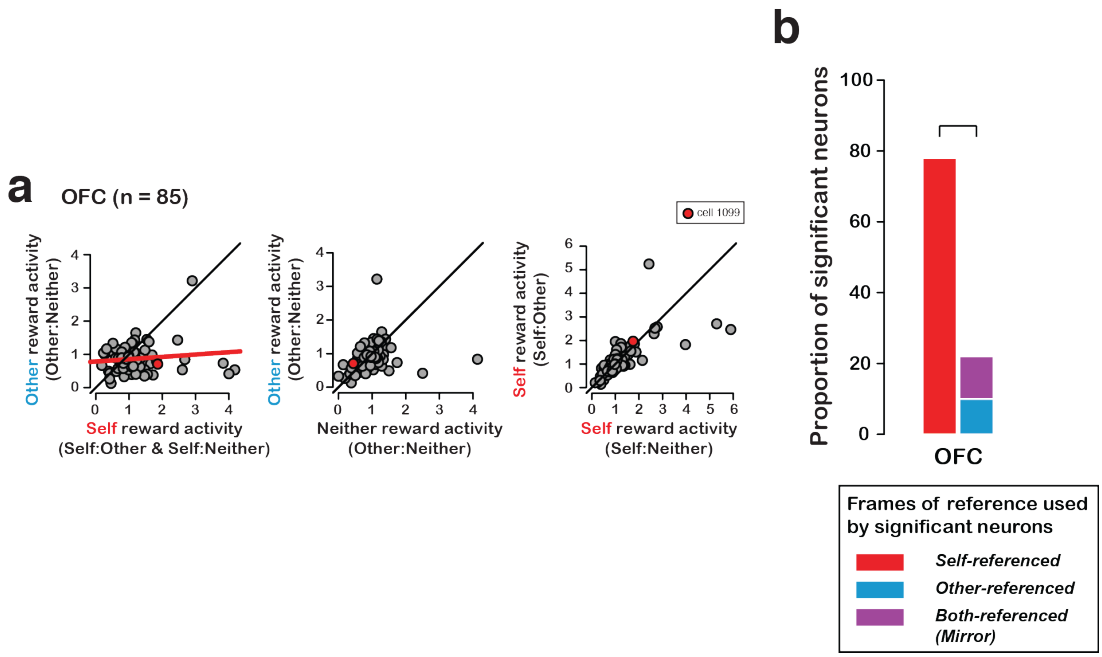


Figure 2. Population biases for *self*, *other*, and *neither* reward outcomes in OFC neurons. (**a**) Scatter plots show mean normalized reward epoch responses of individual neurons between *self* and *other* rewards (left), between *other* and *neither* rewards (middle), and between *self* rewards from *Self:Neither* and *Self:Other* contexts (right). The example neuron from **Fig. 1b** is indicated on the scatter plots. (**b**) Proportion of neurons (out of significantly classified neurons according to ANOVA) from OFC using self-referenced, other-referenced, and both-referenced (mirror) frames for representing reward outcomes. Inset shows color codes used in the bar graph. The asterisk on the bars indicates significant differences in proportions ($P < 0.05$, χ^2 test).

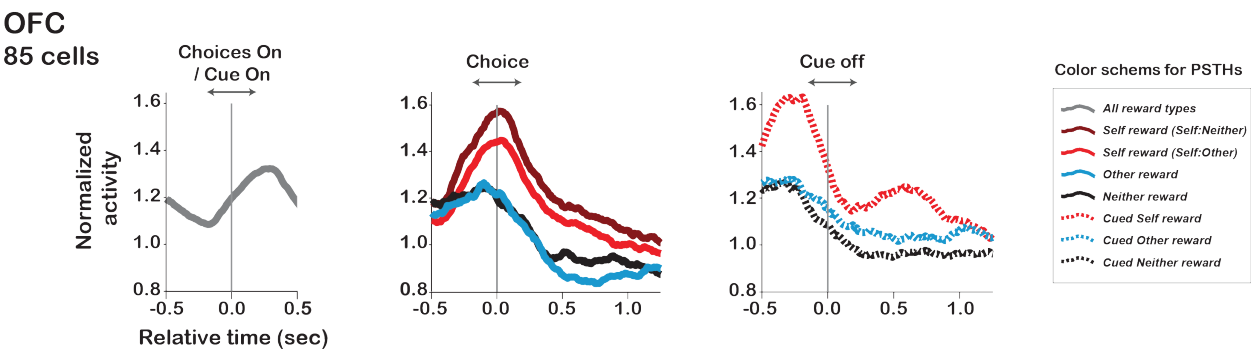


Figure 3 Time courses of the OFC population responses aligned to the time of choice onset and the time of making a choice for choice trials, and aligned to the time of cue onset and cue offset for cued trials. Normalized mean responses (PSTHs) of 85 OFC neurons are plotted over time for each alignment. The inset shows the color scheme. Same format as in **Fig. 1b,c**.

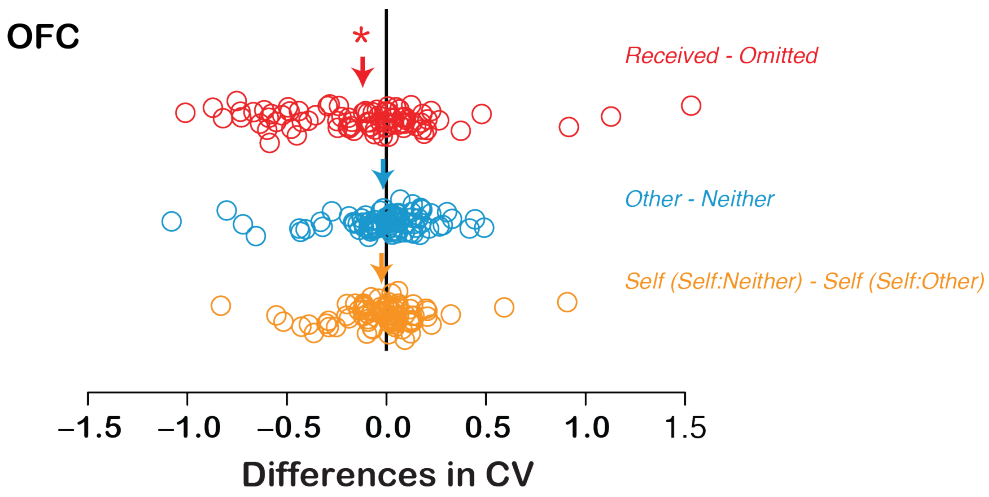


Figure 4 Differences in coefficient of variation (CV) across different reward outcomes also reflect *self* reward bias in OFC. Plotted are differences in CV between a pair of reward categories (as indicated on the right of each distribution). We compared individual neuron averages of all trials in which the donors received rewards against all trials in which the donors did not receive rewards (*Received –Omitted*) (top). We also compared individual neuron averages of trials in which the recipient received the rewards against trials in which no one received rewards (*Other – Neither*) (middle), and, finally, between trials in which the donors received rewards in *Self:Neither* against *Self:Other* contexts (*Self (Self:Neither) – Self (Self:Other)*) (bottom). If applicable, the data were collapsed across choice and cued trials for this analysis. Data points are jittered in the vertical dimensions for visibility. The asterisk above the data points indicates significance (*: $P < 0.10$, one sample *t*-test) in the distribution.

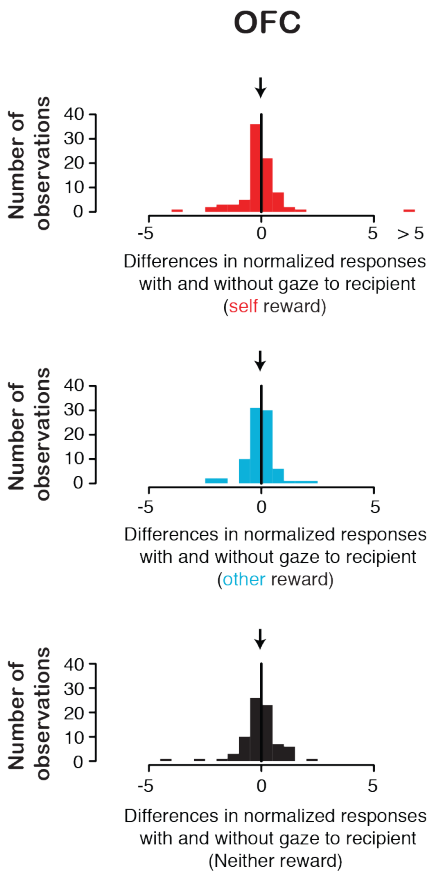


Figure 5 Reward coding is not simply driven by gaze shifts directed at the recipient. Shown are histograms of the differences in normalized reward epoch responses between trials *with* gaze shifts and *without* gaze shifts (responses 'with' – responses 'without' gaze shifts), for trials in which rewards were delivered to *self* (top), *other* (middle), or *neither* (bottom), for OFC populations. Arrows indicate distribution means.

APPENDICES

Appendix 1

Chang SW, Barter JW, Ebitz RB, Watson KK and Platt ML (2012a) Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques (Macaca mulatta). *Proc Natl Acad Sci*, 109, 959–964.

Note: We had begun collecting data for this paper as pilot data for the grant proposal submitted and funded by DoD. We continued to collect data for this paper while we awaited final approval and disbursement of funds from this award. We continued to collect data for this study after disbursement of funds began, and analyzed the data, wrote the paper, and published it while supported by this award.

Appendix 2

Chang SW, Barack DL and Platt ML (2012b) Mechanistic classification of neural circuit dysfunctions: Insights from neuroeconomics research in animals. *Biol. Psychiatry*, 72:101–106.

Appendix 3

Chang SW, Gariépy JF, and Platt ML (*in press)* Neuronal reference frames for social decisions in primate frontal cortex. *Nature Neuroscience..*

Appendix 4

Gariépy JF, Chang SW and Platt ML (*in press*) Brain games: Toward a neuroecology of social behavior. Invited commentary in *Beh. Brain. Sci., in press.*

Statement of Work (SOW)

All monkey experimental sessions will take place in Dr. Michael Platt's dedicated laboratory space in Duke University's Vivarium facility. Data analysis will be conducted in Duke University's Levine Science Research Center.

Platt Laboratory
Center for Cognitive Neuroscience
Duke University
Box 90999
Durham, NC 27708

The laboratory's current approved protocol allows for all experiments proposed in this application, so no period for regulatory review is necessary.

MATLAB scripts for task presentation and behavioral analysis were developed prior to the award period, during the collection of preliminary data.

Timeline:
Proposed start date: 07/01/2011
Proposed end date: 06/30/2014

Pre-Task. USAMRMC ROP IACUC documentation (7/2011 – 10/2011)

Task 1. Characterize neural responses in the orbitofrontal cortex (OFC) (4 monkeys) (10/2011 – 11/2012)

1a. Record cells in OFC while monkeys perform the other-regarding preference (ORP) task (10/2011 – 3/2012)
1b. Analyze recording data (3/2012 – 7/2012)
   a) Characterize activity
   b) Determine response epochs for primary analyses
   c) Correlate neural activity with behavior
1c. Prepare & submit manuscript on OFC recording data (7/2012 – 11/2012)

Task 2. Characterize behavior after muscimol inactivation (4 monkeys) (1/2013 – 9/2013)

2a. Administer muscimol during task performance (1/2013 – 4/2013)
2b. Analyze behavioral data (4/2013 – 6/2013)
   a) Determine the effect of muscimol on behavioral choices
   b) Determine the effect of muscimol on gaze patterns
2c. Prepare & submit manuscript on OFC inactivation (6/2013 – 9/2013)

Task 3. Determine behavioral response to microstimulation (4 monkeys) (9/2013 – 6/2014)

3a. Stimulate OFC during task performance (9/2013 – 2/2014)

3b. Analyze behavioral data (12/2013 – 3/2014)
   a) Determine the effect of microstimulation on behavioral choices
   b) Determine the effect of microstimulation on gaze patterns
   c) Determine the level of perturbation and time-dependency
3c. Prepare & submit manuscript on OFC microstimulation (3/2014 – 6/2014)

Task 4. Examine the effect of oxytocin (OT) on task performance (4 monkeys) (7/2011 – 2/2012)

4a. Deliver OT during task performance (7/2011 – 12/2011)
4b. Analyze behavioral data (9/2011 – 12/2011)
   a) Determine the effect of OT on behavioral choices
   b) Determine the effect of OT on gaze patterns
   c) Examine the time course of the OT effect
4c. Prepare & submit manuscript on OT effect (10/2011 – 2/2012)

Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques (*Macaca mulatta*)

Steve W. C. Chang[a,b,1], Joseph W. Barter[b], R. Becket Ebitz[a,b], Karli K. Watson[a,b], and Michael L. Platt[a,b,c,d]

[a]Department of Neurobiology, Duke University School of Medicine, Durham, NC 27701; and [b]Center for Cognitive Neuroscience and Departments of [c]Evolutionary Anthropology and [d]Psychology and Neuroscience, Duke University, Durham, NC 27708

Edited by Bruce S. McEwen, The Rockefeller University, New York, NY, and approved December 5, 2011 (received for review September 6, 2011)

People attend not only to their own experiences, but also to the experiences of those around them. Such social awareness profoundly influences human behavior by enabling observational learning, as well as by motivating cooperation, charity, empathy, and spite. Oxytocin (OT), a neurosecretory hormone synthesized by hypothalamic neurons in the mammalian brain, can enhance affiliation or boost exclusion in different species in distinct contexts, belying any simple mechanistic neural model. Here we show that inhaled OT penetrates the CNS and subsequently enhances the sensitivity of rhesus macaques to rewards occurring to others as well as themselves. Roughly 2 h after inhaling OT, monkeys increased the frequency of prosocial choices associated with reward to another monkey when the alternative was to reward no one. OT also increased attention to the recipient monkey as well as the time it took to render such a decision. In contrast, within the first 2 h following inhalation, OT increased selfish choices associated with delivery of reward to self over a reward to the other monkey, without affecting attention or decision latency. Despite the differences in species typical social behavior, exogenous, inhaled OT causally promotes social donation behavior in rhesus monkeys, as it does in more egalitarian and monogamous ones, like prairie voles and humans, when there is no perceived cost to self. These findings potentially implicate shared neural mechanisms.

social decision-making | neuropeptide | other-regarding preference | social gaze

Oxytocin (OT) (1) is a mammalian neurosecretory hormone, synthesized by hypothalamic neurons, which regulates the hypothalamic-pituitary-adrenal axis (2). The most well-understood role of OT in mammals is in female reproduction, with peripheral OT influencing parturition and lactation (3), and central OT affecting mother-offspring bonding and recognition (4, 5). More recently, OT has been found to influence nonparental social behavior in a species-specific manner. For example, OT promotes pair-bonding between males and females in monogamous prairie voles (*Microtus ochrogaster*) (6, 7) but can also increase aggression (i.e., mate-guarding behavior) and decrease social interaction among females after brief exposure to a male (8). In humans, OT also influences more complex forms of social behavior and cognition (9–14). For example, inhaled OT enhances trusting behavior toward other individuals in economic games, potentially by suppressing aversion to betrayal risk (15), and promotes cooperation within groups (16). However, inhaled OT also provokes cultural and racial biases (17). OT inhalation also enhances sensitivity to the experiences of others by promoting vicarious reward and empathic pain (10, 18, 19). Recently, OT-mediated processes have been implicated in disorders attended by dysfunctional social behavior, including autism, fragile X syndrome, and schizophrenia (19–22). Notably, OT treatment improves social skills in individuals with autism (21, 23, 24), a spectrum of disorders with marked deficits in sensitivity to what happens to others, including impairments in understanding and responding to social cues (22, 25, 26). Variations in a common oxytocin-receptor allele are linked to autism spectrum disorders and are associated with reduced volume in hypothalamus and anterior cingulate cortex (27).

Despite a growing literature, the mechanisms mediating the influence of OT on sensitivity to what happens to others remain only partially understood (9, 14, 19, 21, 28, 29). OT receptors are localized in multiple regions of the brain, with especially high density in areas implicated in affective and social processing. In prairie voles, OT receptors are densely localized in the amygdala, prelimbic cortex (homologous to the cingulate cortex in primates), and nucleus accumbens of the striatum (30). Recently, it has been shown that OT selectively inhibits a dedicated channel from the central nucleus of the amygdala to periaqueductal gray, ultimately reducing fear-induced freezing behavior in rats (31). Similarly, in humans, inhaled OT influences on social behavior are associated with reduced blood oxygen level-dependent (BOLD) signals in the bilateral amygdala and dorsal striatum (28, 29), consistent with the OT-mediated negative affect processing in the amygdala-cingulate circuits (22). These studies provide evidence that OT influences information processing in neural circuits implicated in emotion and social behavior.

Unlike prairie voles or humans (2, 6, 9–11, 13–16, 30, 32, 33), rhesus macaques (*Macaca mulatta*) live in large, hierarchical social groups with promiscuous mating and uniparental female care of offspring. Precisely how OT might influence social cognition in animals with this type of social structure and mating system, if at all, remains unknown. To answer this question, we capitalized on a recent finding by our group showing that rhesus macaques are sensitive to the rewards experienced by others, and this vicarious reinforcement is sufficient to motivate them to work to reward another monkey when the alternative is delivering reward to no one (34). We found that inhaling OT increased OT levels in cerebral spinal fluid (CSF), demonstrating transnasal penetration into the CNS. Roughly 2 h after OT-inhalation and onward, donor monkeys selectively increased the frequency of choosing an option resulting in reward to an adjacent, visible monkey, when the alternative was rewarding no one. In the same context, OT also increased the frequency that donors looked at the recipient monkey and prolonged choice response times. In contrast, up to about 2 h postinhalation, OT increased selfish decisions when the donors had the option to reward self over the other monkey. These findings invite the hypothesis that OT boosts internal vicarious reinforcement signals in a context-dependent manner in neural circuits homolo-

Author contributions: S.W.C.C. and M.L.P. designed research; S.W.C.C., J.W.B., R.B.E., and K.K.W. performed research; S.W.C.C. analyzed data; S.W.C.C. and M.L.P. wrote the paper; and R.B.E. and K.K.W. oversaw developing and piloting the oxytocin protocol for monkeys.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: steve.chang@duke.edu.

gous to those mediating these processes in humans. Our results demonstrate that OT mediates other-regarding behavior in nonhuman animals, even in those living in despotic societies with uniparental care.

Results

Donor monkeys (hereafter, "self" or "donor") performed a reward allocation task with an unrelated recipient monkey ("other") (Fig. 1 *A*–*C*) (34). The two monkeys were seated in adjacent primate chairs (Crist), 100-cm apart and at 45° angles to each other. Each monkey viewed his own LCD display, and had a juice-tube positioned in front of his mouth through which reward could be delivered. On each trial, donors chose between two visual shapes, associated with rewarding self, other, or neither. We have previously shown that donors typically prefer the shape delivering reward to other over neither (34). This preference is enhanced by greater familiarity between the two monkeys, and is abolished if the recipient monkey is replaced with a juice collection bottle, thus demonstrating the fundamentally social nature of the task (34).

For each session, we intranasally (35) delivered 25 international units (IU) of OT or saline, on alternating days, to two males using a pediatric nebulizer 30 min before performing the reward allocation task. A session composed of multiple reward allocation trials after either OT or saline administration occurred on each day (*Methods*). Data from a total of 12 OT and 10 saline control sessions were collected from two donors (MY and MO) while they engaged in the reward allocation task (Fig. 1 *A*–*C*) with an unrelated recipient monkey (MD). Five OT and three saline sessions were collected from MY, and seven OT and saline sessions each were collected from MO. For statistical power, we present data collapsed across the two donors, unless otherwise stated.

OT inhalation, compared with saline, significantly increased OT concentration in CSF as measured by cervical draws ($P < 0.05$, Welch two-sample t test) (Fig. 1*D*), confirming transnasal penetration into the CNS. Thirty minutes after OT administration, donors began the reward allocation task. For choices between delivering reward to other and neither, OT selectively amplified reward donations to other (Fig. 2). Preference for other increased linearly over time after OT but not after saline (OT: different from 0, $r^2 = 0.26$, $P < 0.0005$; saline: $r^2 = 0.01$, $P = 0.47$, linear regression) (Fig. 2). OT-induced enhancement of prosocial choices was largest in the later half of a given session (i.e., ~110 min after OT administration and ~80 min after task initiation; preference index mean difference between OT vs. saline: 0.17, $P < 0.00001$, Welch two-sample t test) (Fig. 2). Individual donors showed a similar pattern (MY: 0.18, $P < 0.00001$; MO: 0.19, $P < 0.01$). We found a significant difference between the two treatment conditions even when we averaged across the entire duration of the task (mean difference of 0.12, $P < 0.00001$; MY: 0.15, $P < 0.00001$; MO: 0.06, $P < 0.05$, Welch two-sample t test).

In contrast, in the early half of a given session (i.e., up to ~80 min into the task), OT slightly but significantly increased selfish

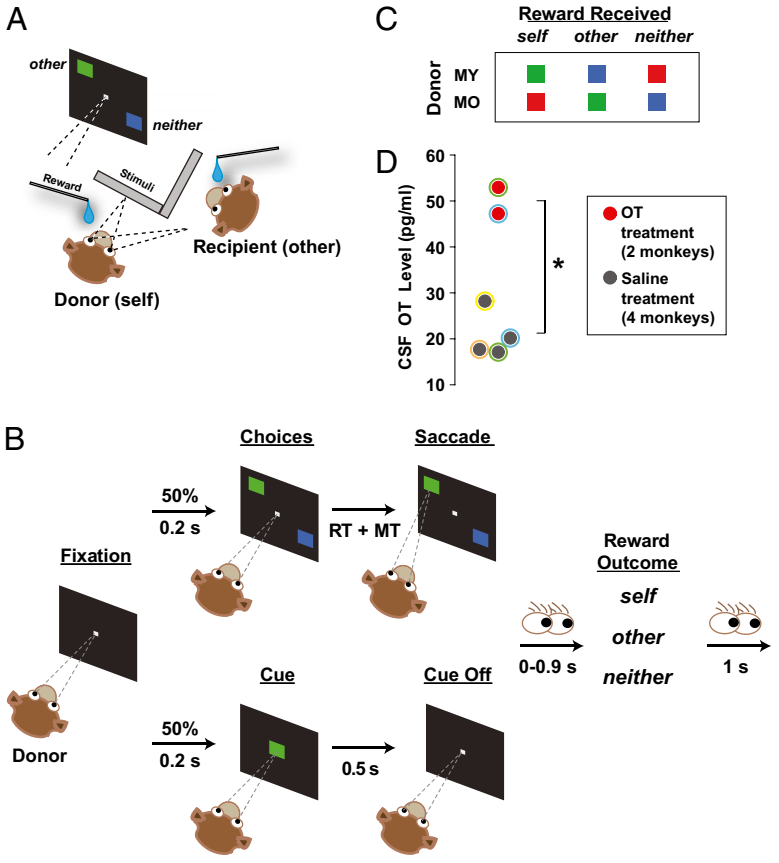


Fig. 1. Reward allocation task. (*A*) Experimental setup. (*B*) Trial sequence. Choice (*Upper*) and cued (*Lower*) trials were randomly interleaved. The eye-gaze cartoons specify the task intervals during which the donors could potentially look at the recipient monkey. MT, movement time; RT, reaction time. (*C*) Stimuli associated with different reward outcomes to donors and recipient, shown separately for the two donors. (*D*) OT concentration in the CSF after intranasal OT (in red) or saline (dark gray). *$P < 0.05$, Welch two-sample t test. Colored outlines on the datapoints represent animal identities.

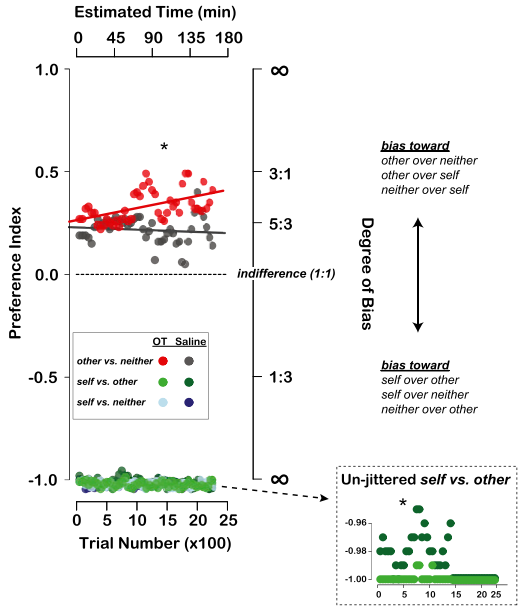


Fig. 2. Intranasal OT promotes both vicarious and self reinforcement. Choice preference index (moving averages of 200 trials per session, 50-trial step) for OT (red) and saline (gray) across all reward options (other vs. neither, self vs. other, and self vs. neither). Datapoints from self vs. other and self vs. neither are jittered along the ordinate for visibility. (*Inset*) Unjittered and magnified data from self vs. other trials. Data from self vs. neither trials were effectively overlapping between the OT and saline conditions, and therefore not shown in an unjittered format. OT, 12 sessions; saline, 10 sessions. Lines show linear regression on other vs. neither trials.

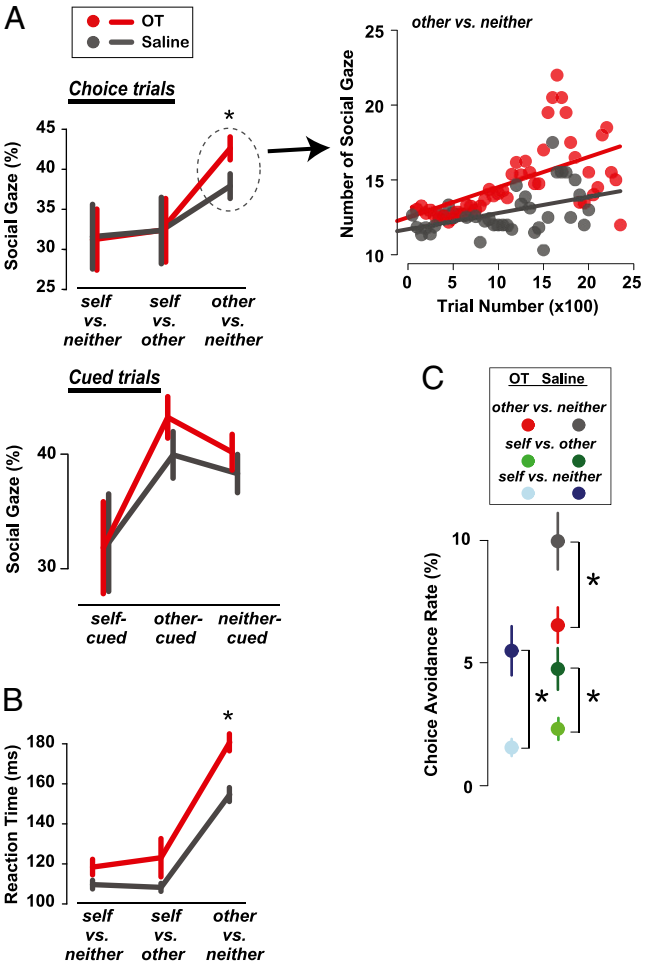


Fig. 3. Intranasal OT enhances attention to the recipient monkey and increases the deliberation time for making donation decisions. (*A*) Gaze to the face of the other monkey after reward delivery. (*Left*) Percentages of gaze shifts to the recipient monkey on choice trials (*Upper*) and cued trials (*Lower*). (*Right*) Number of gaze shifts over the course of each day session for other vs. neither choice trials (moving averages of 200 trials per session, 50-trial step). Lines through the datapoints show linear regressions. (*B*) Response times, measured as saccade onset times following target onset (ms). (*C*) OT reduced choice avoidance [i.e., declining to choose by breaking fixation upon target onset (such as, reward options), which, in the task resulted in a time out for 5 s]. *$P < 0.05$, Welch two-sample t test.

choices on self vs. other trials compared with saline control (mean difference between OT and saline of −0.02, $P < 0.00001$, Welch two-sample t test; *Inset* in Fig. 2 shows unjittered self vs. other trials), but had no effect on self vs. neither trials (mean difference of −0.002, $P = 0.36$). Individual donors showed a similar selfish bias (MY: −0.003, $P < 0.06$; MO: −0.04, $P < 0.00001$). The absence of OT effect on self vs. neither trials might be due to the fact that this context does not involve a potential reward to another monkey, although we cannot rule out the possibility that donors were maximally self-regarding in this context in the absence of OT. Thus, OT robustly enhanced prosocial choices when there was no potential cost to self, but slightly increased selfish choices when there was potential for direct self reward.

Donor monkeys often shift gaze to the recipient monkey after making a choice, and this attention to the recipient is enhanced after prosocial choices compared with selfish choices (34). OT further enhanced this overt other-oriented attention to the recipient after donors made a decision on other vs. neither trials (Fig. 3*A*) (OT vs. saline: mean difference of 4.70%, $P < 0.05$, Welch two-sample t test). In contrast, we did not observe any effects of OT on donor's attention to the recipient when direct self reward was involved (self vs. neither: mean difference of −0.36%, $P = 0.95$; self vs. other: 0.03%, $P = 0.99$) (Fig. 3*A*). We also found that donors looked more frequently to the recipient when rewards were delivered to him compared with when rewards were delivered to self, even on cued trials in which rewards were delivered by computer without any action by donors (gaze frequency on self-cued vs. other-cued trials: OT, $P < 0.005$; saline: $P = 0.05$) (Fig. 3*A*). However, OT did not modulate this difference in social attention on cued trials (all comparisons $P > 0.23$, Welch two-sample t test) (Fig. 3*A*), suggesting that OT enhances other-oriented attention selectively following prosocial decisions rather than in response to anything happening to the other monkey (i.e., after active choices on other vs. neither trials). As in the other-oriented choice preference, attention to the recipient monkey also increased linearly over time after OT (slope significantly different from 0: $r^2 = 0.31$, $P < 0.00001$, linear regression) (Fig. 3*A*, *Right*). The frequency of looking at the recipient monkey in the saline control also increased over the course of the session ($r^2 = 0.19$, $P < 0.005$), but with a significantly lower rate of rise than the OT condition (differences in OT and saline slopes greater than zero: $P < 0.005$, permutation test) (Fig. 3*A*). This finding suggests that OT enhances the intensity of vicarious reinforcement in part by modulating attentional mechanisms.

We also examined the time required by monkeys to render a decision. Response times in the reward allocation task are generally slower when donor monkeys choose between delivering reward to other vs. neither, compared with when self reward is involved (34). OT selectively prolonged response times on other vs. neither trials (mean difference between OT and saline of 26.0 ms, $P < 0.00001$, Welch two-sample t test) (Fig. 3*B*), possibly

reflecting internal processes, such as deliberation and control. On self vs. neither and self vs. other trials, however, OT only showed a trend on response times (self vs. other: mean difference of 14.78 ms; self vs. neither: 8.72 ms; both $P < 0.13$) (Fig. 3*B*). Finally, on some trials, donors avoided making a decision, opting to wait until the next trial (although they could not predict the subsequent reward options). OT reduced this choice avoidance behavior across all trial types (all $P < 0.05$, Welch two-sample t test) (Fig. 3*C*), perhaps because of overall enhancement in subjective reinforcement.

Inhaled OT thus influenced reward donation decisions by rhesus macaques when there was an option to reward another monkey (other vs. neither and self vs. other, but not self vs. neither). OT enhanced reward donations on other vs. neither trials, but increased selfish behavior on self vs. other trials (Fig. 2). OT-induced changes in attention to the recipient monkey (Fig. 3*A*) and decision time (Fig. 3*B*) were both specific to the donation context (other vs. neither), whereas OT-induced reductions in choice avoidance behavior (Fig. 3*C*) were global.

Discussion

Compared with some other nonhuman primates, social behavior of rhesus monkeys is primarily characterized by competition and aggression, and shows very weak, if any, inclination toward cooperation (36, 37). In a prior study, different levels of endogenous OT were reported in more socially affiliative mother-reared compared with more socially agnostic nursery-reared macaques (38). Here we show that exogenous OT promotes social donation behavior in rhesus macaques, as it does in more egalitarian and monogamous species, like prairie voles and humans. OT-induced prosocial donations were accompanied by enhanced other-oriented attention and decision times. In contrast, in a context in which there was a potential for rewarding self or another monkey, OT slightly increased the tendency for donors to choose selfishly without influencing overt attention and, at most, minimally affecting decision times. The absence of OT-induced enhancement of overt attention on these trials suggests that OT modulates other-oriented preferences through vicarious reinforcement (34). These findings are consistent with context-dependent effects of OT on human social behavior (16, 17, 39) (for a review of human social processing, see ref. 40), implying similar neural mechanisms.

Given the context-specific increase in attention to the other monkey and more deliberative decision latency, it is conceivable that these behaviors are related. Several hypotheses are plausible. On the one hand, OT may increase attention to the other monkey via neural circuits mediating orienting behavior, including amygdala, parietal cortex, and superior colliculus. Increased attention to the recipient may enhance vicarious reinforcement experienced from delivering juice to him. Alternatively, OT may influence neural circuits involved in decision-making, including the striatum and anterior cingulate cortex (see introductory paragraphs). Slowed response times may reflect more deliberate processing of the potential outcomes available (41). A future study designed to probe the temporal evolution of OT-induced effects on attention and decision-making will be needed to resolve these hypotheses.

The direction of OT-induced social enhancement also appears to vary as a function of time. OT initially enhanced self reinforcement but later amplified vicarious reinforcement, although the largest OT-induced effects were prosocial. Although this interaction between time-dependent and context-dependent effects of OT may be specific to our reward allocation task and thus can only be extrapolated with caution, these results suggest that OT may influence self- and other-regarding behaviors via distinct underlying neural mechanisms.

Why might OT promote self reinforcement bias on self vs. other but not on self vs. neither trials? The key difference between the two contexts is the alternative option. In one context, the alternative option has a social consequence (i.e., rewarding the recipient), whereas in the other context, the alternative option does not (i.e., nothing happens to either donor or recipient). OT-induced self reinforcement may depend on the contrast between rewarding self and another individual. We hypothesize that when a decision context presents this contrast, OT can promote selfish behavior. OT influences on self and vicarious reinforcement (16, 17, 39) thus appear to depend on the social state of the underlying neural circuits.

Previous studies in monogamous prairie voles and promiscuous montane voles (*Microtus montanus*) have suggested that mating system may be a key predictor of OT influences on social behavior through the topology of OT receptor localization in neural circuits, mediating reinforcement and motivation (33). A more general difference between prairie voles and montane voles is the frequency and intensity of social interaction (33). Compared with montane voles, prairie voles are biparental, show more selective aggression, and spend more time in close physical proximity (33). Humans and rhesus macaques, too, are highly social mammals; intranasal OT induces prosocial tendencies in humans (15, 16) and, as we now report, in rhesus macaques. These findings suggest that OT may play a critical role in modulating social behavior in highly gregarious mammals, regardless of mating system or parental care strategy.

Intranasal administration of OT in humans has also been shown to increase gaze to the eyes of others (19). We found that OT enhanced gaze directed at the face of the other monkey following active social decision-making but not following passive reward delivery. This finding invites the possibility that OT gates the activity of attention circuits in the brain specifically during active interaction with others. Evidence from human functional neuroimaging studies is consistent with this idea. For example, OT selectively modulates BOLD signal in the anterior cingulate cortex, amygdala, midbrain, and dorsal striatum during a trust game involving other human players, but not during a nonsocial decision-making task (29). Functional connectivity between the amygdala and midbrain structures is also reduced by OT when human participants view emotional faces (28). Finally, OT reduces the subjective evaluation of aversively conditioned faces, and this reduction is accompanied by suppressed BOLD responses in the amygdala and the fusiform gyrus (42).

Consistent with our results, OT modulates deliberation times during social decision-making in humans. For example, OT slows overall evaluation time for rating faces in a nonspecific manner, regardless of whether the images were aversively conditioned or not (42). OT can also speed up decision times; for example, OT decreased overall key press reaction times for evaluating in-group favoritism and out-group derogation in an implicit association test (17).

OT enhanced the frequency of prosocial decisions in the absence of opportunity for direct self reward, but provoked an increase in selfish decisions when choosing between self and other. Such a dual function has also been reported in humans. OT can both promote cooperation and increase out-group bias depending on behavioral context (16, 17, 39). Thus, OT does not appear to have a universal prosocial influence on behavior, but rather amplifies ongoing social information processing (21), perhaps by influencing already existing preferences. It is plausible that OT mediates prosociality and generosity only in an indirect manner. Alternatively, OT may play a more direct and causal role in modulating context-dependent social information processing (e.g., refs. 27–29 for neural evidence), specifically by enhancing the gain of neural circuits mediating vicarious reinforcement and attention.

Recently, OT has been evaluated for potential therapeutic use in clinical conditions attended by dysfunctional social behavior, such as autism spectrum disorders, antisocial personality disorder, and schizophrenia (20–24, 43, 44). Notably, the intranasal nebulization method (35) we developed here is well-tolerated by children for

delivery of other therapeutics (i.e., albuterol), thus opening up avenues for early OT intervention in neuropsychiatric conditions with social deficits. Furthermore, choice-specific effect of OT on increasing other-oriented attention suggests a potential need for active decision-making during OT interventions.

The current finding opens up new opportunities for uncovering the mechanisms underlying the influences of OT on social behavior in a species much more closely related to humans than rodents. Rhesus monkeys have long served as the primary model species for probing the neural mechanisms mediating high-level cognition. Given the strong similarities in social behavior and cognition, and the apparent homologies in underlying neural circuitry, the rhesus macaque provides a powerful model for probing the mechanisms mediating some of the basic behaviors that make complex human social interactions possible.

Methods

General Procedures and Behavioral Task. All procedures were approved by the Duke University Institutional Animal Care and Use Committee. Two donor monkeys (MY and MO) and a recipient monkey (MD) participated in the study. All animals underwent standard surgical procedures for implanting a head-restraint prosthesis at least 6 mo before the present study. The head-restraint prosthesis allowed us to monitor eye position, sampled at 1,000 Hz (SR Research; Eyelink), as well as conduct single-unit recordings in other experiments, not reported here. Both the donor and recipient were head-restrained throughout the experiment. Donors and recipient were unrelated, middle-ranked, and not cage mates. Face of recipient (other; corresponding horizontal and vertical eye positions) was empirically mapped. Rewards were 0.5–1.0 mL of cherry-flavored juice. Within each block, reward size was constant for all three outcomes. A separate solenoid was designated for rewarding neither that only produced clicks but delivered no fluid. To prevent monkeys from forming secondary associations between solenoid clicks and different reward types, all solenoid valves (including the one used to deliver "neither" reward) used to deliver juice rewards were placed in another room. Masking white noise was also played in the experimental room.

Donors began the trial by shifting gaze (± 2.5°) to a central stimulus (0.5° × 0.5°), and maintained fixation (for 200 ms). Choice and cue trials were presented at equal frequencies and randomly interleaved. On choice trials (Fig. 1*B*), two visual targets (4° × 4°) appeared at two random locations of 7° eccentricity and reflected about the vertical meridian. Donors shifted their gaze to one target (± 2.5°) to indicate their choice. On cued trials (Fig. 1*B*), donors maintained fixation while a cue appeared centrally (for 500 ms). On both trial types, the reward onset was followed by a 0 to 0.9 s delay. Donors could freely look around for 0–0.9 s following making a choice and for another 1 s after the reward onset. Data from error trials are not included in analyses.

Data from 12 OT (MY: 5, MO: 7) and 10 saline (MY: 3, MO: 7) sessions were collected on strictly alternating days. Each day session was, on average, 1,274 ± 141 (mean ± SEM) trials. Within each day session, several blocks of the task (a median of 6 and 6.5 blocks for OT and saline, respectively) were completed by the donors. Each of these blocks typically consisted of 192 ± 10 (mean ± SEM) and 205 ± 15 trials for OT and saline, respectively.

Intranasal OT Protocol. Donor monkeys were transported in the primate chair from the colony room to the experimental room. After stabilizing their heads, OT (25 IU/mL; Agrilabs) was delivered via nebulization (Pari Baby Nebulizer) into the nose and mouth continuously for 5 min (5 IU/min) when the donor monkeys were fully awake. On alternating days, nebulized saline served as a control. Before experimental sessions, donor monkeys were first habituated to the nebulizer and then accustomed to saline delivery using the nebulizer in an incremental fashion until they were completely relaxed during the procedure, which typically took about a week. In fact, donor monkeys showed no distress during this procedure. Testing began exactly 30 min after each treatment, at which time a recipient monkey was brought to the experimental setup. In the guinea pig CNS, radioactively labeled OT lasts up to 4 h (45). In humans, intranasal delivery of a similar peptide, vasopressin (differing by only two amino acids), increases its concentration in the CSF after 10 min, and elevated vasopressin levels are maintained for more than 80 min after administration (35). In that study (35), vasopressin levels increased significantly after 30 min. Previous studies in humans have not measured inhaled OT uptake into the CNS. Fig. 1*D* plots CSF OT levels in monkeys 35 min after inhalation, demonstrating efficacy of the intranasal nebulization method (see below). Note that the mask was always pressed very tightly to minimize potential leakage, but nonetheless leakage could have occurred. It is worth noting that CSF OT levels may have continued to increase after the time of CSF measurement, warranting caution in linking absolute CSF OT levels with changes in behavior. Despite these uncertainties, our nebulization technique resulted in a ~2.5-fold increase in CSF OT levels roughly 0.5 h after inhalation.

CSF OT Protocol. To determine whether inhaled OT penetrates the CNS after nebulization, OT concentration in CSF was measured via cervical punctures (on average 35 min after the beginning of inhalation). Cervical punctures were performed by a licensed veterinarian, and targeted the cisterna magna through the juncture between the occipital base and atlas (C1) through the atlanto-occipital membrane. Monkeys were first anesthetized with ketamine (3 mg/kg, i.m.) and dexdomitor (0.075 mg/kg, i.m.). To reverse anesthesia, we administered antisedan (0.075 mg/kg, i.m.) once the animal was returned to its cage after the draw. Approximately 0.5 mL of CSF was drawn using a 24 to 27 gauge needle. At the performing veterinarian's discretion, bupivacaine was administered subcutaneously at the insertion site following needle removal. CSF was immediately frozen on dry ice and sent off-site to be assayed for OT (Biomarkers Core Labs, Yerkes National Primate Research Center, Atlanta, GA) using a commercially prepared kit [Assay Designs (now Enzo Life Sciences); cat. # 900–153: Oxytocin ELISA kit, with very low reactivity with vasopressin]. Samples were assayed "neat" with a range of 15.6–1,000 μL assay volume. This assay has near-zero reactivity with vasopressin, which is chemically similar to OT, thus providing specific quantitation of OT.

Data Analysis. Preference index was a contrast ratio of frequency of choosing an option, n_A or n_B:

$$\textit{Preference Index} = \frac{n_A - n_B}{n_A + n_B}.$$

For choices between self vs. other, n_A and n_B were number of choices to reward other and self, respectively. For choices between other vs. neither, n_A and n_B were number of choices to reward other and neither, respectively. Finally, for choices between self vs. neither, n_A and n_B were number of choices to reward neither and self, respectively. Indices ranged from –1 to 1, with 1 corresponding to always choosing the "prosocial" option to reward the recipient monkey (when that was an option) or to withhold reward from self (self vs. neither). An index of –1 indicated that donors always chose an "antisocial" option to reward self (when that was an option) or to withhold reward from the other monkey (other vs. neither). Preference index of 0 indicated indifference. Frequency of donors looking at recipients was computed from number of gaze shifts to the recipient's facial region (within ± 8.5° spanning from the center of the recipient's face). Reaction times (time from target onset to movement onset) were computed using a 20°/s velocity threshold (46).

ACKNOWLEDGMENTS. We thank Ernst Fehr, Markus Heinrichs, and Steven P. Wise for helpful discussions and valuable suggestions; Benjamin Y. Hayden, David B. Barack, and Jean-Francois Gariépy for comments; Monica L. Carlson for technical assistance; and Francis Sun DVM at the Duke Division of Laboratory Animal Resource for cerebrospinal fluid collection. This work was supported by a Ruth K. Broad Biomedical Foundation Postdoctoral grant (to S.W.C.C.); a Predoctoral grant (to R.B.E.); the Davis Foundation (K.K.W.), a National Institutes of Health T32 Postdoctoral Grant (to S.W.C.C.); and National Institutes of Health Grant MH086712 (to S.W.C.C., K.K.W., and M.L.P.).

1. Du Vigneaud V, et al. (1953) The synthesis of an octapeptide amide with the hormonal activity of oxytocin. *J Am Chem Soc* 75:4879–4880.
2. Donaldson ZR, Young LJ (2008) Oxytocin, vasopressin, and the neurogenetics of sociality. *Science* 322:900–904.
3. Soloff MS, Alexandrova M, Fernstrom MJ (1979) Oxytocin receptors: Triggers for parturition and lactation? *Science* 204:1313–1315.
4. Pedersen CA (1999) *Oxytocin Control of Maternal Behavior: Regulation of Sex Steroids and Offspring Stimuli. The Integrative Neurobiology of Affiliation*, eds Carter CS, Lederhendler I, Kirkpatrick B (MIT Press, Cambridge, MA), pp 301–320.
5. Feldman R, Weller A, Zagoory-Sharon O, Levine A (2007) Evidence for a neuroendocrinological foundation of human affiliation: Plasma oxytocin levels across pregnancy and the postpartum period predict mother-infant bonding. *Psychol Sci* 18:965–970.
6. Lim MM, Young LJ (2006) Neuropeptidergic regulation of affiliative behavior and social bonding in animals. *Horm Behav* 50:506–517.
7. Bales KL, et al. (2007) Oxytocin has dose-dependent developmental effects on pair-bonding and alloparental care in female prairie voles. *Horm Behav* 52:274–279.
8. Bales KL, Carter CS (2003) Sex differences and developmental effects of oxytocin on aggression and social behavior in prairie voles (*Microtus ochrogaster*). *Horm Behav* 44:178–184.

9. Fehr E, Camerer CF (2007) Social neuroeconomics: The neural circuitry of social preferences. *Trends Cogn Sci* 11:419–427.
10. Singer T, et al. (2008) Effects of oxytocin and prosocial behavior on brain responses to direct and vicariously experienced pain. *Emotion* 8:781–791.
11. Stein DJ (2009) Oxytocin and vasopressin: Social neuropeptides. *CNS Spectr* 14: 602–606.
12. Heinrichs M, von Dawans B, Domes G (2009) Oxytocin, vasopressin, and human social behavior. *Front Neuroendocrinol* 30:548–557.
13. Zak PJ, Kurzban R, Matzner WT (2005) Oxytocin is associated with human trustworthiness. *Horm Behav* 48:522–527.
14. Meyer-Lindenberg A (2008) Impact of prosocial neuropeptides on human brain function. *Prog Brain Res* 170:463–470.
15. Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005) Oxytocin increases trust in humans. *Nature* 435:673–676.
16. De Dreu CK, et al. (2010) The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* 328:1408–1411.
17. De Dreu CK, Greer LL, Van Kleef GA, Shalvi S, Handgraaf MJ (2011) Oxytocin promotes human ethnocentrism. *Proc Natl Acad Sci USA* 108:1262–1266.
18. Zak PJ, Stanton AA, Ahmadi S (2007) Oxytocin increases generosity in humans. *PLoS ONE* 2:e1128.
19. Guastella AJ, Mitchell PB, Dadds MR (2008) Oxytocin increases gaze to the eye region of human faces. *Biol Psychiatry* 63:3–5.
20. Bartz JA, Hollander E (2006) The neuroscience of affiliation: Forging links between basic and clinical research on neuropeptides and social behavior. *Horm Behav* 50: 518–528.
21. Insel TR (2010) The challenge of translation in social neuroscience: A review of oxytocin, vasopressin, and affiliative behavior. *Neuron* 65:768–779.
22. Meyer-Lindenberg A, Domes G, Kirsch P, Heinrichs M (2011) Oxytocin and vasopressin in the human brain: Social neuropeptides for translational medicine. *Nat Rev Neurosci* 12:524–538.
23. Andari E, et al. (2010) Promoting social behavior with oxytocin in high-functioning autism spectrum disorders. *Proc Natl Acad Sci USA* 107:4389–4394.
24. Hollander E, et al. (2007) Oxytocin increases retention of social cognition in autism. *Biol Psychiatry* 61:498–503.
25. Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a "theory of mind"? *Cognition* 21:37–46.
26. Insel TR, Fernald RD (2004) How the brain processes social information: Searching for the social brain. *Annu Rev Neurosci* 27:697–722.
27. Tost H, et al. (2010) A common allele in the oxytocin receptor gene (OXTR) impacts prosocial temperament and human hypothalamic-limbic structure and function. *Proc Natl Acad Sci USA* 107:13936–13941.
28. Kirsch P, et al. (2005) Oxytocin modulates neural circuitry for social cognition and fear in humans. *J Neurosci* 25:11489–11493.
29. Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E (2008) Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* 58: 639–650.
30. Insel TR, Shapiro LE (1992) Oxytocin receptor distribution reflects social organization in monogamous and polygamous voles. *Proc Natl Acad Sci USA* 89:5981–5985.
31. Viviani D, et al. (2011) Oxytocin selectively gates fear responses through distinct outputs from the central amygdala. *Science* 333:104–107.
32. Bales KL, Pfeifer LA, Carter CS (2004) Sex differences and developmental effects of manipulations of oxytocin on alloparenting and anxiety in prairie voles. *Dev Psychobiol* 44:123–131.
33. Young LJ, Wang Z, Insel TR (1998) Neuroendocrine bases of monogamy. *Trends Neurosci* 21:71–75.
34. Chang SW, Winecoff AA, Platt ML (2011) Vicarious reinforcement in rhesus macaques (*Macaca mulatta*). *Front Neurosci* 5:27.
35. Born J, et al. (2002) Sniffing neuropeptides: A transnasal approach to the human brain. *Nat Neurosci* 5:514–516.
36. Maestripieri D (2007) *Macachiavellian Intelligence: How Rhesus Macaques and Humans Have Conquered the World* (Univ of Chicago Press, Chicago).
37. Thierry B, Mewa S, Kaumanns W (2004) *Macaque Societies: A Model for the Study of Social Organization* (Cambridge University Press, Cambridge; New York, NY) xv, 418 pp.
38. Winslow JT, Noble PL, Lyons CK, Sterk SM, Insel TR (2003) Rearing effects on cerebrospinal fluid oxytocin concentration and social buffering in rhesus monkeys. *Neuropsychopharmacology* 28:910–918.
39. Bartz JA, Zaki J, Bolger N, Ochsner KN (2011) Social effects of oxytocin in humans: Context and person matter. *Trends Cogn Sci* 15:301–309.
40. Fehr E, Fischbacher U (2003) The nature of human altruism. *Nature* 425:785–791.
41. Gold JI, Shadlen MN (2001) Neural computations that underlie decisions about sensory stimuli. *Trends Cogn Sci* 5:10–16.
42. Petrovic P, Kalisch R, Singer T, Dolan RJ (2008) Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *J Neurosci* 28:6607–6615.
43. Domes G, Heinrichs M, Michel A, Berger C, Herpertz SC (2007) Oxytocin improves "mind-reading" in humans. *Biol Psychiatry* 61:731–733.
44. Bartz JA, Hollander E (2008) Oxytocin and experimental therapeutics in autism spectrum disorders. *Prog Brain Res* 170:451–462.
45. Robinson IC, Jones PM (1982) Oxytocin and neurophysin in plasma and CSF during suckling in the guinea-pig. *Neuroendocrinology* 34:59–63.
46. Paré M, Munoz DP (1996) Saccadic reaction time in the monkey: Advanced preparation of oculomotor programs is primarily responsible for express saccade occurrence. *J Neurophysiol* 76:3666–3681.

Mechanistic Classification of Neural Circuit Dysfunctions: Insights from Neuroeconomics Research in Animals

Steve W.C. Chang, David L. Barack, and Michael L. Platt

Many psychiatric conditions present complex behavioral symptoms, and the type and magnitude of underlying neural dysfunction may vary drastically. This review introduces a classification scheme for psychiatric symptoms, describing them in terms of the state of a dysfunctional neural circuit. We provide examples of two kinds of functional deficits: variance-shifted functionality, in which a damaged circuit continues to function albeit suboptimally, and state-shifted functionality, resulting in an absent or qualitatively different functional state. We discuss, from the perspective of neuroeconomics and related areas of behavioral investigation, three broad classes of commonly occurring symptoms in psychopathology based on selected studies of decision making in animals: temporal discounting, social preferences, and decision making under environmental volatility. We conclude that the proposed mechanistic categorization scheme offers promise for understanding neural circuit dysfunctions underlying psychopathology.

Key Words: Animals, decision, electronic circuit, neuroeconomics, psychopathology, reward, state-shifted, suboptimal, variance-shifted

Comprised of constellations of behavioral symptoms, psychiatric disorders frequently frustrate any simple attempt to translate observed phenotype into neurobiological mechanism. Even at the individual symptom level, such translation is challenging and not easily quantifiable. Behavioral symptoms are often compound and thus difficult to interpret. This presents a challenge for understanding their core neurobiological features, creating practical barriers to designing behavioral or diagnostic tests. This difficulty may be amplified when studying the illnesses manifested as a result of dysfunctions in the prefrontal, limbic, and paralimbic regions, which are less well understood, compared with, for example, the occipital cortex. A promising alternative to understanding the neurobiology of psychiatric disorders begins by classifying them according to the ways the underlying mechanisms may fail. In this issue exploring the benefits of a neuroeconomics approach for understanding psychopathology, we outline a mechanistic classification scheme grounded in the principles of neuroeconomic studies of cognition and behavior in animals.

Variance-Shifted Versus State-Shifted Functionality: Insights from Electronics

Dysfunctional neural circuitry can be functionally classified into two different states based on the outputs of disrupted circuits. As an illustration, consider an electronic circuit designed to produce a specific output. A variance-shifted circuit operates with added noise and, therefore, generates a broadened output distribution, resulting in suboptimal performance. However, a suboptimal circuit may continue to process information (1). By contrast, a state-shifted circuit may generate a completely different functional output, either beyond the expectation of a downstream circuit or failing to generate any output at all, producing a qualitatively different or absent output and resulting in behavior drawn from a different distribution altogether (1).

As a simplified analogy, a simple band-pass filter illustrates the different classes of damage-induced functional states. A change in circuit resistance or capacitance will change the effective cutoff frequency, while a short in the system effectively halts filtering (1). Changes in a circuit's resistance will result in a noisier output, analogous to psychiatric conditions in which afflicted individuals show difficulty in evaluating changes in the environment. Such damage to the circuit reveals its critical role for producing adaptive, normal behavior. In contrast, the presence of a short in the system will prevent filtering of relevant information, analogous to situations where afflicted individuals completely lose sensitivity to changes in the environment. In this case, the state-shifted circuit reveals its necessary role in the production of a particular behavior.

The intricate balance between circuit components can result in functional changes that are either large and noticeable or small and subtle. Some neuropsychiatric symptoms only differ from others slightly, whereas others are so specific to a condition that they serve as a diagnostic hallmark. Furthermore, because of the complex and multilayered nature of neural circuits, initial perturbations may result at first in a state-shifted circuit that, due to neural plasticity, resolves back to a variance-shifted, or even fully restored, state. In summary, psychiatric symptoms may result from a relatively preserved neural circuit operating with added noise, producing deviant and suboptimal behavior (variance-shifted functionality). Alternatively, it may arise from a shorted circuit producing completely different or absent behaviors (state-shifted functionality).

The two damaged states can be described in terms of neural network models as well. In a trained neural network, the organizational principles involve individual computational units, or nodes, whose functionalities may be obscure and may encode information idiosyncratically (2,3). A variance-shifted functional state may result from damage to peripheral nodes, whereas a state-shifted state may be induced by damage to a central node in the network. The two functionalities can also be described based on the output statistics of an implicated circuit. A variance-shifted dysfunction in a neural circuit may produce circuit (or behavioral) outputs characterized by a broadened and/or attenuated distribution compared with optimal functionality (thus less specific or more noisy). In contrast, a state-shifted dysfunction in a circuit may produce an output

From the Department of Neurobiology (SWCC, MLP), Duke University School of Medicine; and Center for Cognitive Neuroscience (SWCC, DLB, MLP) and Departments of Philosophy (DLB), Evolutionary Anthropology (MLP), and Psychology and Neuroscience (MLP), Duke University, Durham, North Carolina.

Address correspondence to Michael L. Platt, Ph.D., Duke University, Department of Neurobiology, Center for Cognitive Neuroscience, LSRC Building Room B243F, Durham, NC 27710; E-mail: platt@neuro.duke.edu.

Received Aug 25, 2011; revised Feb 20, 2012; accepted Feb 21, 2012.

0006-3223/$36.00
doi:10.1016/j.biopsych.2012.02.017

drawn from a completely different distribution (thus qualitatively different) or may result in a complete failure to produce any output. It is worthwhile to note that a state shift could occur in the direction of extreme enhancement, resulting in exaggerated behavior such as positive symptoms in schizophrenia.

Our classification scheme, though neither exceptionless nor exhaustive, provides insight into the possible mechanisms underlying psychiatric symptoms. The two deficit types may occur simultaneously or sequentially (and the distinction sometimes can be ambiguous until a given circuit is fully understood) but may provide novel mechanistic insights into psychopathology and inform the relationship of pathology to health. This approach differs fundamentally from the *Diagnostic and Statistical Manual of Mental Disorders*, the *International Classification of Diseases*, and the like, which are designed to describe a disorder using a list of behavioral symptoms for diagnostic purposes. The present scheme is useful for directly comparing the functionality of neural mechanisms and their corresponding behaviors across normal and dysfunctional states of the brain. A successful distinction between variance- and state-shifted dysfunction is constrained by our understanding of a given circuit. For example, a variance-shifted dysfunction under one functional criterion could be seen as a state-shifted condition under a different framework. Such ambiguity, which is present in any classification scheme, can only be resolved through more comprehensive understanding of a circuit.

Examples from Oculomotor System

Examples from oculomotor system help illustrate the two distinct dysfunctional states described above. The superior colliculus and frontal eye fields belong to a distributed oculomotor circuit spanning cortical and subcortical structures (4,5). Frontal eye field lesions increase variability in saccade trajectories and severely disrupt selection of targets in the contralesional hemifield (6). Frontal eye field lesioned animals, however, can still saccade (6). By contrast, superior colliculus lesions temporarily abolish contralesional saccades altogether (7). They also permanently increase saccade latencies and eliminate the animal's ability to make express saccades (saccades with reaction times less than 100 msec in monkeys) in a gap task (7), designed to bypass the time required to disengage from visual fixation by inserting a gap between the offset of a fixation stimulus and target onset (8). Therefore, for saccades, frontal eye field disruption results in noisy (i.e., variable) performance but preserves overall functionality, a variance-shifted dysfunction. Superior colliculus damage alone, by contrast, is sufficient to temporarily abolish saccades, which is consistent with a state-shifted dysfunction. These examples demonstrate that distinct mechanistic deficits can impair or abolish normal function.

Neuroeconomics of Decision Making in Animals

Neuroeconomics, a discipline that marries the mathematical formalisms of classical economics, the psychophysical methods of behavioral economics, and contemporary neurosciences (9–11), provides an illuminating test of the functionality-based classification scheme for defining mechanistic pathologies in decision making (for a review regarding the benefits of animal models in neuroeconomics, see [12]). The approach applies mathematically tractable economic formalizations to the nervous system and focuses on basic economic concepts such as utility (9,13–15), risk (16,17), and temporal discounting (18,19), providing quantitative frameworks for examining the neural mechanisms underlying cognitive processes (12).

The neuroeconomic framework in animal models is advantageous for studying complex forms of decision making by tapping into their innate reward-seeking behaviors while maintaining ethological validity. Unlike in humans, animal models offer access to studying complex behaviors at the resolution of single neurons. Further, insights into different types of mechanistic deficits in neuropsychiatric symptoms can be obtained by studying decisions animals make following perturbation of neural circuits. Thus, animal models of decision making provide valuable insights into characterizing the biological mechanisms of behavior, detailing the formal operations the brain performs in realizing different cognitive capacities.

We discuss a selection of experiments, categorizing the observed deficits as the variance-shifted and state-shifted model of neural circuit dysfunctions. We organize this discussion around three examples of circuit dysfunction in light of neuroeconomics and other related disciplines: disorders of temporal discounting in addiction, social and other-regarding preferences (ORP), and decision making under environmental volatility. Our intention is not to establish necessary and sufficient conditions for connecting a specific dysfunction and a specific neural circuit. Doing so would not be practically possible. Instead, in this exercise, we attempt to label experimentally induced behavioral deficits observed in animals as dysfunctions arising from either a variance- or state-shifted functional state in the implicated circuit. Although this classification scheme can be just as easily applied to any perturbation results (e.g., microstimulation or drug infusion), we focus on lesion studies for their blunt effectiveness in perturbing circuit function.

Addiction as a Disorder of Temporal Discounting

Single-unit recordings in animals, as well as neuroimaging in humans, have found that striatal dopaminergic signaling is critical for reward-related processing, including motivation and learning (20–22), and that dysfunctional dopaminergic signaling disrupts reward anticipation in drug addiction (for a review, see [23–25]). Firing rates of midbrain dopamine neurons compute economic decision parameters, such as reward probability, reward delay, and reward uncertainty (26–28). Dopaminergic signaling is also involved in evaluating the economic costs and benefits of upcoming rewards. For example, neurons in rodent nucleus accumbens (NAc) encode anticipated reward benefits, without encoding response costs to achieve the reward (28). Such economic computations by the mesolimbic dopamine system may contribute to addiction and other motivation-related disorders.

Temporal discounting describes a time-dependent devaluation of economic value (18). It is a phenomenon observed across multiple species including rodents, monkeys, and humans (18,29,30). When provided an option to choose an immediate but smaller reward over a larger reward with a longer delay, animals reliably prefer the immediate option (31). Addicted individuals discount more than nonaddicted individuals (24,32), as evidenced by behaviors manifested in addiction to cocaine, alcohol, opioid, nicotine, and gambling (for a review, see [32]). Therefore, a disruption in temporal discounting may be a common mechanistic deficit shared by many classes of addiction.

Single-unit recordings in monkeys demonstrate that neurons in the striatum mediate computations underlying temporal discounting (33). Rats with NAc lesions display severe difficulty in choosing a delayed reward option in an intertemporal choice task, suggesting a critical role of NAc in computing economic values of rewards in time (34). Further, NAc lesions do not abolish reward sensitivity altogether but impair the implementation of an optimal (reward-

maximizing) strategy (35), as if these animals cannot accurately compute temporally discounted utility to guide decisions. Similarly, addicted individuals rarely lose the ability to seek addicted substances. Rather, they display impaired impulsive control in pursuing immediate rewards, consistent with atypical temporal discounting. Thus, addiction-related deficits resemble a variance-shifted functionality, resulting in disrupted decisions in time, though retaining some sensitivity to reward (i.e., performance does not become random and the discounting function does not become flat). Deficits resulting from perturbations to dopamine circuits performing economic calculations seem to cause noisy mappings, or variance shifts in the representations, among reward, action, and time.

Neural correlates of temporal discounting are also found in the prefrontal cortex (for a review, see [36]). Neurons in dorsolateral prefrontal cortex (dlPFC) encode the temporally discounted value of upcoming rewards (19). A cocaine self-administration study in monkeys found that activity in the anterior cingulate cortex (ACC) is enhanced upon cocaine intake (37), consistent with human neuroimaging studies showing that drug seeking in addiction is linked to the prefrontal cortex (38,39). ACC involvement in reward-guided decision making is not limited to processing directly experienced outcomes but also includes fictive outcomes (40), similar to the human ventral striatum (41). Correctly utilizing such fictive signals may be critical in addiction. Individuals with chronic nicotine addiction fail to utilize these signals to adjust their choices in an investment task (42). Furthermore, gambling addiction seems to require rewards that are delivered according to a partial or a variable schedule (43), coupled with near-miss fictive reward signals.

Disorders of Social and Other-Regarding Preferences

Precisely how social information is integrated into economic decisions in neural circuits remains obscure. Understanding whether social disorders are manifested by a deficit in a decision circuit or a circuit purely involved in evaluating social information from the environment remains a challenge. ORPs describe a consideration for the economic well-being of others. ORP computations may reflect a stage where decision making and social information processing are partially integrated. Consider autism spectrum disorder (ASD), which handicaps social and communicative abilities of ~1 per 110 children in the United States (44). Autism spectrum disorder individuals show little interest in others (45). This lack of interest is associated with other complex social deficits, including reduced empathy and joint attention, further disrupting the capacity for normal social interactions (46,47). Differences between ASD and typically developing individuals are illustrated by performance in economic bargaining games designed to elicit ORP. While healthy individuals readily engage in reciprocal cooperation in these games, ASD individuals adopt simple rules that are both less flexible and more laboriously employed (48). It remains unclear whether circuit dysfunctions in ASD more closely resemble variance-shifted or state-shifted states. Comparison with other disorders marked by social deficits, such as schizophrenia, psychopathy, and eating disorders, may help to illuminate the underlying pathology in ASD.

ACC is critical for social processing. ACC gyrus lesions in monkeys abolish the animal's ability to evaluate social information, as measured by response latencies to retrieve food in the presence of socially arousing images, such as staring monkeys (49). Although the changes in response latencies in ACC-lesioned animals can differ substantially depending on the types of social stimuli and often on the individuals, sensitivity to social stimuli can be eliminated by the lesion (49). This social evaluation deficit therefore resembles a state-shifted functionality, in which social evaluation processing is no longer intact. In contrast, ACC sulcus and orbitofrontal cortex (OFC) lesions produce deviant behaviors but fail to abolish the sensitivity to social stimuli (49), resembling a noisy suboptimal state and a variance-shifted functionality.

Closely related to ORP, empathy-related processing by ACC has been investigated in the context of perceiving painful events of others. The brain areas involved in pain perception in humans, namely ACC and frontal insula, are more metabolically active when perceiving a painful stimulus delivered to fair compared with unfair players in an economic game (50). In rodents, ACC, along with other medial pain systems, mediates observational fear conditioning while watching a conspecific receive a shock (51). Both lidocaine-induced inactivation and targeted deletion of a voltage-gated calcium channel in ACC can substantially reduce observational fear conditioning but not eliminate it (51). A dysfunction in empathy-related processing in ACC might be driven by variance-shifted dysfunctional states, resulting in degraded sensitivities to process or simulate the painful events of others.

A link between ORP and emotional processing remains elusive. Amygdala is one of the primary structures linked to emotional processing and is reciprocally connected to ACC and OFC (52,53). Amygdala dysfunction is related to a number of psychiatric symptoms, including major depression and bipolar disorder and affective psychosis in schizophrenia (54). Typically, amygdala contribution to emotional processing has been investigated using fear-inducing or social stimuli. Monkeys with bilateral amygdala lesions show abolished fear responses, as measured by response latencies to retrieve food in the presence of a fearful stimulus (52,55). Consistent with these observations, amygdala-lesioned rats completely lose the ability to acquire conditioned fear, even when the lesion occurs a month after the initial Pavlovian training, suggesting a necessary role in emotional memory (56,57) (i.e., state-shifted due to an absent distribution). Notably, in many psychiatric conditions involving emotion, the gain on emotional processing in amygdala might be set too high, possibly due to impaired communication with other structures, such as prefrontal cortex, that modulate amygdala activity (58). Such unregulated emotional processing might lead to exaggerated behavior, presumably due to a state shift. For example, this state shift might result in a more responsive and less regulated state. The reciprocal information transmission among the amygdala, ACC, and OFC (52,53) suggests that the emotional component of ORP may originate from the amygdala.

Disorders of Decision Making Under Environmental Volatility

Several neurological and psychiatric disorders compromise the adaptive abilities of cognitive systems, whether updating the expected values of targets according to task demands or appropriately reorienting to reflect changes in the environment. Notably, some cognitive deficits such as an inflexibility to adapt to environmental changes are shared across multiple neurological and psychiatric conditions. For example, degeneration of mechanisms that contribute to adaptive decision making, including task set switching, task set maintenance, and inhibitory control, characterizes cognitive and executive deficits in schizophrenia (59,60). From a neuroeconomic perspective, these may emerge from failures in updating reward valuation, risk, and volatility. In the Wisconsin Card Sorting Task (WCST), typically used to probe the ability to adjust to changing environments without explicit cues, participants sort a deck of cards according to unpredictably changing rules (61). During the task, schizophrenic patients perseverate more on choosing

incorrect responses, persisting longer with a previous rule despite negative feedback (62). These individuals also show increased response times and make more errors in the Stroop task (63–67).

Schizophrenia is accompanied by both negative symptoms, such as lack of emotion, and positive symptoms, such as hallucinations and delusions (68). In addition, schizophrenia is associated with deficits in executive and cognitive functions (68). Such deficits include inflexible adjustments in behavioral strategies, or policies, that require computing expected value of reinforcers on the basis of the accumulation of evidence over time, assessment of value on the basis of reinforcer identity, and projecting these evaluations into the future (69). Schizophrenic patients also show decreased abilities to stay on task (70,71). Deficits related to executive control are suggested to be caused by noisy dopaminergic gating of prefrontal neurons (70). Symptoms in the domain of executive control may thus reflect variance-shifted processing. Positive and negative symptoms, on the other hand, are associated with exaggerated (e.g., hallucinations) and abolished (e.g., lack of emotion) processing, respectively, and thus are more consistent with a state-shifted condition.

In schizophrenia, the posterior cingulate cortex (PCC) is associated with increased default network connectivity, with the degree of enhanced connectivity positively correlating with the severity of psychopathology, and these patients show increased cannabinoid receptor expression (mediating inhibitory neurotransmitters like gamma-aminobutyric acid) (72,73). A case study of lesions in the human PCC found an inability to adapt to new environments (74). Consistent with this, neuronal activity in monkey PCC tracks the level of risk in changing environments (17) and is correlated with setting a behavioral strategy to explore or exploit different options (75,76). Thus, disruptions to PCC seem to compromise an ability to detect and incorporate discontinuities in environmental statistics such as changes in expected value and risk. It remains unclear whether volatility-related deficits in PCC lesions reflect variance- or state-shifted functionalities.

An explicit task-switching paradigm, in which a correct response on a given trial or group of successive trials is explicitly cued, is often used to investigate executive control. In such a task, neurons in ACC increase responses following task switches (77), suggesting sensitivity to changes in reward information used in executive control. Lesions to ACC gyrus increase the frequency of consecutive errors, whereas more comprehensive lesions in ACC (gyrus and sulcus) result in slowed response times, errors in switching, and greater overall consecutive errors (78). Critically, although ACC lesions increase switch-related errors, monkeys are still able to switch tasks above the chance level, suggesting the mechanisms responsible for cognitive flexibility are not completely abolished (78). These results implicate a variance-shifted deficit inducing suboptimality in the ability to adapt to changing environments by explicit changes in the expected values of the targets.

Perseveration of maladaptive behavior is one of the most striking features of prefrontal lesions. Such deficits are apparent in environments without explicit rule-changing cues. In WCST, patients with dlPFC lesions fail to switch to a correct response and instead perseverate on an incorrect response (79). Indeed, schizophrenia is associated with inefficient dlPFC function, particularly with respect to working memory (80). Activity of dlPFC neurons in monkeys is correlated with the level of conflict in WCST (81) and different strategies employed within the task (82,83). In a WCST analog, lesions to monkey OFC, ACC, or dlPFC in and around the principal sulcus (but not superior and medial to the sulcus) all result in fewer uncued rule-guided behavioral shifts, though the animals still execute switches, indicating variance-shifted, as opposed to fully state-shifted, dysfunction (84). In contrast, dlPFC-lesioned animals no longer show a stereotypical increase in response times as a function of conflict, an abolition of conflict-induced changes in motor responses (81), consistent with the full destruction of conflict-detection mechanisms in dlPFC (state-shifted). Conflict detection and resolution in these tasks may map onto running calculations of instantaneous utility and uncertainty, though this remains a topic of ongoing debate. By perturbing circuits that detect conflict or encode strategy, dlPFC damage leads to a computational deficiency in value updating for flexible environmental adaptation.

Conclusions

We are just beginning to understand what constitutes a psychiatric disease. Neuroeconomic studies in animals provide new insights into the affected neural circuits (85). Our proposed classification scheme establishes a new framework for thinking about psychiatric disorders formulated in the language of neural circuits. It remains to be seen how the circuit-based classification could augment the existing typological schemes to help assess and treat psychiatric disorders. As a first step, we have focused on deficits tied to specific breakdowns in selected neural circuits. Some deficits are shared and thus might appear in multiple classically defined illnesses. Our interpretation is intended to point out that what superficially might appear to be very different syndromes may, in fact, share common disruptions in the underlying neural circuitry.

Psychopathological symptoms can be approached based on the precise type of deficits induced in neural circuits. A neural circuit will show different outputs depending on the affected circuit components. A noisy state broadens the width of the output distribution, leading to suboptimal performance, but may not alter the basic functionality of a given circuit. In contrast, a circuit could break down or be extensively modified, introducing a new state into the system with abnormal or absent functionalities that are qualitatively different from the norm.

Most psychiatric disorders present compound symptoms. It is not surprising then that a single psychiatric illness arises from a combination of variance-shifted and state-shifted circuit dysfunctions, involving multiple brain areas. For example, under a connectionist neural network framework, variance-shifted dysfunctions may result from damages to peripheral processing nodes. When the most critical region of the distributed network is disrupted, however, we may observe a fully compromised, state-shifted dysfunction instead (though the deficits may eventually be restored by other areas in the network on a longer time scale). Note that there are clear cases of state-shifted psychopathology when the deficits are not due to targeted traumatic brain injury. For example, in visual or auditory hallucinations, commonly found with severe schizophrenia, individuals experience percepts in the absence of actual sensory signals. The circuits that mediate these experiences are clearly behaving very differently and seem likely to be induced by a state-shifted process.

Our circuit-based scheme may be relevant for the ongoing debate in psychiatry over the need for incorporating dimensional diagnosis to traditional categorical diagnosis (86–90). The variance- and state-shifted models effectively redescribe such dimensional criteria at the level of neural circuits. For example, the severity or idiosyncrasy of a given symptom for a given individual could be linked to either the degree of variance shift (e.g., the magnitude of change in the variance of the distribution) or the degree of state shifts (e.g., the magnitude of mean shifts in the distribution) in behavioral or cognitive output according to the proposed scheme. Translating psychiatric symptoms into dimensional outcomes of

neural circuit dysfunction may open up new avenues for improved therapeutic intervention.

The circuit-based classification does not describe a relationship between implicated circuits and psychiatric disorder types. Our classification scheme, which critically depends on our understanding of the functionality of a given circuit, is not intended to replace existing typologies of psychopathology. Rather, it describes a mechanistic relationship between implicated circuits and behavioral deficits caused by failures of those circuits. In our view, the current scheme can provide easily quantifiable grounds for hypothesis testing for linking a circuit-level dysfunction and an afflicted behavior (e.g., Supplement 1) and thus may provide novel insights into the mechanistic dysfunctions underlying psychiatric conditions.

This work is supported by National Institutes of Health 5T32NS051156-07 (SWCC), National Institute of Mental Health 5R01MH086712-03 (DLB and MLP), and Department of Defense AR100035 (SWCC and MLP).

We are grateful to Nancy L. Zucker and Geoffrey K. Adams for helpful feedback.

All authors declare no biomedical financial interests or potential conflicts of interest.

Supplementary material cited in this article is available online.

1. Horowitz P, Hill W (1989): The Art of Electronics, *2nd ed.* Cambridge, England: Cambridge University Press.
2. Poggio T, Edelman S (1990): A network that learns to recognize three-dimensional objects. *Nature* 343:263–266.
3. Poggio T (1990): A theory of how the brain might work. *Cold Spring Harb Symp Quant Biol* 55:899–910.
4. Corbetta M, Akbudak E, Conturo TE, Snyder AZ, Ollinger JM, Drury HA, *et al.* (1998): A common network of functional areas for attention and eye movements. *Neuron* 21:761–773.
5. Ferraina S, Pare M, Wurtz RH (2002): Comparison of cortico-cortical and cortico-collicular signals for the generation of saccadic eye movements. *J Neurophysiol* 87:845–858.
6. Schiller PH, Chou IH (1998): The effects of frontal eye field and dorsomedial frontal cortex lesions on visually guided eye movements. *Nat Neurosci* 1:248–253.
7. Schiller PH, Sandell JH, Maunsell JH (1987): The effect of frontal eye field and superior colliculus lesions on saccadic latencies in the rhesus monkey. *J Neurophysiol* 57:1033–1049.
8. Pare M, Munoz DP (1996): Saccadic reaction time in the monkey: Advanced preparation of oculomotor programs is primarily responsible for express saccade occurrence. *J Neurophysiol* 76:3666–3681.
9. Glimcher PW (2009): *Neuroeconomics: Decision Making and the Brain*. London: Academic Press.
10. Loewenstein G, Rick S, Cohen JD (2008): Neuroeconomics. *Annu Rev Psychol* 59:647–672.
11. Camerer CF (2008): Neuroeconomics: Opening the gray box. *Neuron* 60:416–419.
12. van Wingerden M, Kalenscher T (2011): Why we should use animals to study economic decision making–a perspective. *Front Neurosci* 5:82.
13. Platt ML, Glimcher PW (1999): Neural correlates of decision variables in parietal cortex. *Nature* 400:233–238.
14. Montague PR, Berns GS (2002): Neural economics and the biological substrates of valuation. *Neuron* 36:265–284.
15. Kable JW, Glimcher PW (2007): The neural correlates of subjective value during intertemporal choice. *Nat Neurosci* 10:1625–1633.
16. Platt ML, Huettel SA (2008): Risky business: The neuroeconomics of decision making under uncertainty. *Nat Neurosci* 11:398–403.
17. McCoy AN, Platt ML (2005): Risk-sensitive neurons in macaque posterior cingulate cortex. *Nat Neurosci* 8:1220–1227.
18. Green L, Myerson J (2004): A discounting framework for choice with delayed and probabilistic rewards. *Psychol Bull* 130:769–792.
19. Kim S, Hwang J, Lee D (2008): Prefrontal coding of temporally discounted values during intertemporal choice. *Neuron* 59:161–172.
20. Schultz W, Dayan P, Montague PR (1997): A neural substrate of prediction and reward. *Science* 275:1593–1599.
21. Montague PR, Hyman SE, Cohen JD (2004): Computational roles for dopamine in behavioural control. *Nature* 431:760–767.
22. McLaren I (1989): The computational unit as an assembly of neurones: An implementation of an error correcting learning algorithm. In: Durbin R, Miall C, Mitchison G, editors. *The Computing Neuron*. Amsterdam: Addison-Wesley, 160–178.
23. Hyman SE, Malenka RC, Nestler EJ (2006): Neural mechanisms of addiction: The role of reward-related learning and memory. *Annu Rev Neurosci* 29:565–598.
24. Schultz W (2011): Potential vulnerabilities of neuronal reward, risk, and decision mechanisms to addictive drugs. *Neuron* 69:603–617.
25. Wise RA (1996): Neurobiology of addiction. *Curr Opin Neurobiol* 6:243–251.
26. Fiorillo CD, Tobler PN, Schultz W (2003): Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299:1898–1902.
27. Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H (2006): Midbrain dopamine neurons encode decisions for future action. *Nat Neurosci* 9:1057–1063.
28. Gan JO, Walton ME, Phillips PE (2010): Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. *Nat Neurosci* 13:25–27.
29. Hwang J, Kim S, Lee D (2009): Temporal discounting and inter-temporal choice in rhesus monkeys. *Front Behav Neurosci* 3:9.
30. Hayden BY, Platt ML (2007): Temporal discounting predicts risk sensitivity in rhesus macaques. *Curr Biol* 17:49–53.
31. Myerson J, Green L (1995): Discounting of delayed rewards: Models of individual choice. *J Exp Anal Behav* 64:263–276.
32. Bickel WK, Miller ML, Yi R, Kowal BP, Lindquist DM, Pitcock JA (2007): Behavioral and neuroeconomics of drug addiction: Competing neural systems and temporal discounting processes. *Drug Alcohol Depend* 90(suppl 1):S85–S91.
33. Cai X, Kim S, Lee D (2011): Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron* 69:170–182.
34. Cardinal RN, Pennicott DR, Sugathapala CL, Robbins TW, Everitt BJ (2001): Impulsive choice induced in rats by lesions of the nucleus accumbens core. *Science* 292:2499–2501.
35. Parkinson JA, Olmstead MC, Burns LH, Robbins TW, Everitt BJ (1999): Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. *J Neurosci* 19:2401–2411.
36. Kim S, Lee D (2011): Prefrontal cortex and impulsive decision making. *Biol Psychiatry* 69:1140–1146.
37. Baeg EH, Jackson ME, Jedema HP, Bradberry CW (2009): Orbitofrontal and anterior cingulate cortex neurons selectively process cocaine-associated environmental cues in the rhesus monkey. *J Neurosci* 29:11619–11627.
38. Garavan H, Pankiewicz J, Bloom A, Cho JK, Sperry L, Ross TJ, *et al.* (2000): Cue-induced cocaine craving: Neuroanatomical specificity for drug users and drug stimuli. *Am J Psychiatry* 157:1789–1798.
39. Goldstein RZ, Alia-Klein N, Tomasi D, Carrillo JH, Maloney T, Woicik PA, *et al.* (2009): Anterior cingulate cortex hypoactivations to an emotionally salient task in cocaine addiction. *Proc Natl Acad Sci U S A* 106:9453–9458.
40. Hayden BY, Pearson JM, Platt ML (2009): Fictive reward signals in the anterior cingulate cortex. *Science* 324:948–950.
41. Lohrenz T, McCabe K, Camerer CF, Montague PR (2007): Neural signature of fictive learning signals in a sequential investment task. *Proc Natl Acad Sci U S A* 104:9493–9498.
42. Chiu PH, Lohrenz TM, Montague PR (2008): Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. *Nat Neurosci* 11:514–520.
43. Sharpe L, Tarrier N (1993): Towards a cognitive-behavioural theory of problem gambling. *Br J Psychiatry* 162:407–412.
44. Rice C (2009): Prevalence of autism spectrum disorders–Autism and Developmental Disabilities Monitoring Network, United States, 2006. *MMWR Surveill Summ* 58:1–20.
45. Kanner L (1943): Autistic disturbances of affective contact. *Nervous Child* 2:217–250.

46. Batson C, Duncan B, Ackerman P, Buckley T, Birch K (1981): Is empathic emotion a source of altruistic motivation? *J Pers Soc Psychol* 40:290–302.
47. Goldman A (1993): Ethics and cognitive science. *Ethics* 103:337–360.
48. Sally D, Hill E (2006): The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. *J Econ Psychol* 27:73–97.
49. Rudebeck PH, Buckley MJ, Walton ME, Rushworth MF (2006): A role for the macaque anterior cingulate gyrus in social valuation. *Science* 313: 1310–1312.
50. Singer T, Seymour B, O'Doherty JP, Stephan KE, Dolan RJ, Frith CD (2006): Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439:466–469.
51. Jeon D, Kim S, Chetana M, Jo D, Ruley HE, Lin S-Y, *et al.* (2010): Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. *Nat Neurosci* 13:482–488.
52. Murray EA (2007): The amygdala, reward and emotion. *Trends Cogn Sci* 11:489–497.
53. Phelps EA, LeDoux JE (2005): Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron* 48:175–187.
54. Krishnamoorthy ES (2007): A differential role for the hippocampus and amygdala in neuropsychiatric disorders. *J Neurol Neurosurg Psychiatry* 78:1165–1166.
55. Izquierdo A, Murray EA (2007): Selective bilateral amygdala lesions in rhesus monkeys fail to disrupt object reversal learning. *J Neurosci* 27: 1054–1062.
56. Fanselow MS, LeDoux JE (1999): Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron* 23:229–232.
57. Maren S, Aharonov G, Fanselow MS (1996): Retrograde abolition of conditional fear after excitotoxic lesions in the basolateral amygdala of rats: Absence of a temporal gradient. *Behav Neurosci* 110:718–726.
58. Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002): Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav Rev* 26:321–352.
59. Kerns JG, Nuechterlein KH, Braver TS, Barch DM (2008): Executive functioning component mechanisms and schizophrenia. *Biol Psychiatry* 64: 26–33.
60. Meiran N, Levine J, Henik A (2000): Task set switching in schizophrenia. *Neuropsychology* 14:471–482.
61. Monchi O, Petrides M, Petre V, Worsley K, Dagher A (2001): Wisconsin Card Sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *J Neurosci* 21:7733–7741.
62. Everett J, Lavoie K, Gagnon JF, Gosselin N (2001): Performance of patients with schizophrenia on the Wisconsin Card Sorting Test (WCST). *J Psychiatry Neurosci* 26:123–130.
63. Barch DM, Carter CS, Perlstein W, Baird J, Cohen JD, Schooler N (1999): Increased stroop facilitation effects in schizophrenia are not due to increased automatic spreading activation. *Schizophr Res* 39:51–64.
64. Henik A, Salo R (2004): Schizophrenia and the stroop effect. *Behav Cogn Neurosci Rev* 3:42–59.
65. Crider A (1997): Perseveration in schizophrenia. *Schizophr Bull* 23:63–74.
66. McNeely HE, West R, Christensen BK, Alain C (2003): Neurophysiological evidence for disturbances of conflict processing in patients with schizophrenia. *J Abnorm Psychol* 112:679–688.
67. Koren D, Seidman LJ, Harrison RH, Lyons MJ, Kremen WS, Caplan B, *et al.* (1998): Factor structure of the Wisconsin Card Sorting Test: Dimensions of deficit in schizophrenia. *Neuropsychology* 12:289–302.
68. Barch DM (2005): The cognitive neuroscience of schizophrenia. *Annu Rev Clin Psychol* 1:321–353.
69. Daw ND, Niv Y, Dayan P (2005): Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8:1704–1711.
70. Braver TS, Barch DM, Cohen JD (1999): Cognition and control in schizophrenia: A computational model of dopamine and prefrontal function. *Biol Psychiatry* 46:312–328.
71. Kieffaber PD, Kappenman ES, Bodkins M, Shekhar A, O'Donnell BF, Hetrick WP (2006): Switch and maintenance of task set in schizophrenia. *Schizophr Res* 84:345–358.
72. Newell KA, Deng C, Huang XF (2006): Increased cannabinoid receptor density in the posterior cingulate cortex in schizophrenia. *Exp Brain Res* 172:556–560.
73. Whitfield-Gabrieli S, Thermenos HW, Milanovic S, Tsuang MT, Faraone SV, McCarley RW, *et al.* (2009): Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proc Natl Acad Sci U S A* 106:1279–1284.
74. Katayama K, Takahashi N, Ogawara K, Hattori T (1999): Pure topographical disorientation due to right posterior cingulate lesion. *Cortex* 35: 279–282.
75. Pearson JM, Heilbronner SR, Barack DL, Hayden BY, Platt ML (2011): Posterior cingulate cortex: Adapting behavior to a changing world. *Trends Cogn Sci* 15:143–151.
76. Hayden BY, Nair AC, McCoy AN, Platt ML (2008): Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron* 60:19–25.
77. Johnston K, Levin HM, Koval MJ, Everling S (2007): Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. *Neuron* 53:453–462.
78. Rushworth MF, Hadland KA, Gaffan D, Passingham RE (2003): The effect of cingulate cortex lesions on task switching and working memory. *J Cogn Neurosci* 15:338–353.
79. Stuss DT, Levine B, Alexander MP, Hong J, Palumbo C, Hamer L, *et al.* (2000): Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia* 38:388–402.
80. Potkin SG, Turner JA, Brown GG, McCarthy G, Greve DN, Glover GH, *et al.* (2009): Working memory and DLPFC inefficiency in schizophrenia: The FBIRN study. *Schizophr Bull* 35:19–31.
81. Mansouri FA, Buckley MJ, Tanaka K (2007): Mnemonic function of the dorsolateral prefrontal cortex in conflict-induced behavioral adjustment. *Science* 318:987–990.
82. Genovesio A, Brasted PJ, Mitz AR, Wise SP (2005): Prefrontal cortex activity related to abstract response strategies. *Neuron* 47:307–320.
83. Tsujimoto S, Genovesio A, Wise SP (2011): Comparison of strategy signals in the dorsolateral and orbital prefrontal cortex. *J Neurosci* 31:4583–4592.
84. Buckley MJ, Mansouri FA, Hoda H, Mahboubi M, Browning PG, Kwok SC, *et al.* (2009): Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science* 325:52–58.
85. Kishida KT, King-Casas B, Montague PR (2010): Neuroeconomic approaches to mental disorders. *Neuron* 67:543–554.
86. Pickles A, Angold A (2003): Natural categories or fundamental dimensions: On carving nature at the joints and the rearticulation of psychopathology. *Dev Psychopathol* 15:529–551.
87. Krueger RF, Watson D, Barlow DH (2005): Introduction to the special section: Toward a dimensionally based taxonomy of psychopathology. *J Abnorm Psychol* 114:491–493.
88. Goldberg D (2000): Plato versus Aristotle: Categorical and dimensional models for common mental disorders. *Compr Psychiatry* 41:8–13.
89. Haslam N (2003): Categorical versus dimensional models of mental disorder: The taxometric evidence. *Aust N Z J Psychiatry* 37:696–704.
90. Helzer JE, Kraemer HC, Krueger RF (2006): The feasibility and need for dimensional psychiatric diagnoses. *Psychol Med* 36:1671–1680.

Mechanistic Classification of Neural Circuit Dysfunctions: Insights from Neuroeconomics Research in Animals

Supplemental Information

An Example of Hypothesis Testing Under the Variance- and State-Shifted Framework

Lesions to the dorsolateral prefrontal cortex (dlPFC) in monkeys impair executive control (1). The posterior cingulate cortex (PCC) has been implicated in tracking volatility in the environment (2). Both dlPFC and PCC are implicated in schizophrenia (see Main Text). However, it remains unknown how the two areas interact to exert flexible cognitive control. Recording neuronal activity from PCC after a dlPFC lesion could provide a unique opportunity to test how the executive control impairments due to a dlPFC lesion affect volatility-tracking signals in PCC. More precisely, comparing the response profiles of PCC neurons before and after the dlPFC lesion could reveal whether the prefrontal lesion induces either variance-shifted or state-shifted dysfunctions in PCC neurons. Similarly, a neuroimaging experiment could test, in individuals with schizophrenia who show abnormal dlPFC and PCC metabolic activity, whether and how (e.g., variance- or state-shifted) the activations in dlPFC and PCC are functionally linked. Results from studies like these can help reveal novel insights into how certain circuits malfunction during specific behaviors being tested. Unlike traditional classification schemes of psychiatric symptoms, the current circuit-based scheme can provide straightforward testable grounds for understanding how a given circuit dysfunction might be related to behavioral deficits observed in psychiatric conditions.

Supplemental References

1. Buckley MJ, Mansouri FA, Hoda H, Mahboubi M, Browning PG, Kwok SC, *et al.* (2009): Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science*. 325:52-58.
2. McCoy AN, Platt ML (2005): Risk-sensitive neurons in macaque posterior cingulate cortex. *Nat Neurosci*. 8:1220-1227.

Manuscript type
Article

Title
Neuronal reference frames for social decisions in primate frontal cortex

Authors
Steve W. C. Chang [1,2]*, Jean-Francois Gariépy [2], and Michael L. Platt [1,2,3]

Running title
Neuronal reference frames for social decisions

Affiliations
[1] Department of Neurobiology,
Duke University School of Medicine, Durham, NC 27710
[2] Center for Cognitive Neuroscience,
[3] Department of Psychology and Neurosciences,
Department of Evolutionary Anthropology,
Duke Institute for Brain Sciences,
Duke University, Durham, NC 27708

Manuscript Information
Main Content: Main Text, Online Methods, 7 Figures
Number of total references: 52
Supplementary Information Content: 1 Table, 9 Figures

Manuscript Correspondence *
Steve W.C. Chang, Ph.D.
Department of Neurobiology
Center for Cognitive Neuroscience
B203 Levine Science Research Center
Duke University
Box 90999
Durham, NC 27708, USA
Tel: (919)-668-0333
Fax: (919)-681-0815
Email: steve.chang@duke.edu

Abstract

Social decisions play a crucial role in the success of individuals and the groups they compose. Group members respond vicariously to benefits obtained by others, and impairments in this capacity contribute to neuropsychiatric disorders like autism and sociopathy. We studied how neurons in three frontal cortical areas encode the outcomes of social decisions as monkeys performed a reward-allocation task. Neurons in the orbitofrontal cortex (OFC) predominantly encoded rewards delivered to oneself. Neurons in the anterior cingulate gyrus (ACCg) encoded reward allocations to the other monkey, reward allocations to oneself, or both. Neurons in the anterior cingulate sulcus (ACCs) signaled reward allocations to the other monkey or no one. Within this network of received (OFC) and foregone (ACCs) reward signaling, ACCg emerges as a key nexus for the computation of shared experience and social reward. Individual and species-specific variations in social decision-making might result from the relative activation and influence of these areas.

Social cohesion depends on vicarious identification with members of one's group. In social situations, we are aware of our actions and their consequences but also consider those of others, especially those with whom we might interact[1]. We also estimate the internal states of others, perhaps by simulation[2], which in turn shapes our future actions. Social situations can drive observational learning[3], and other-regarding preferences influence neural computations that ultimately result in cooperation, altruism, or spite[4,5]. Disruptions of neural circuits involved in other-regarding processes may underlie social deficits attending neuropsychiatric conditions like autism[6]. Human imaging and clinical studies have found critical links between social deficits and abnormal brain activity in frontal cortex and its subcortical targets[7].

Neural circuits involved in reinforcement learning and decision-making are crucial for normal social interactions[8]. Critical nodes include the anterior cingulate cortex[9–11], the orbitofrontal cortex[12–17], and subcortical areas such as the dopaminergic ventral tegmental area and substantia nigra[18,19], the striatum[20–21], the lateral habenula[22], and the amygdala[23]. Neuroimaging studies in humans report activation of some of these areas by both giving rewards and receiving rewards[24–28], and lesions to some of these areas result in impaired social decision-making[7]. These findings thus suggest a generic circuit for reward-guided learning and decision-making mediates social decisions[8]. Despite this evidence, and the clear clinical relevance of understanding the neurobiology of social decision-making, precisely how neurons in any of these areas compute social decisions remains unknown, largely due to difficulties in implementing social interactions while simultaneously studying neuronal activity and controlling contextual variables. Single unit recording studies in nonhuman animals, such as macaques, making social decisions of similar complexity to those made by humans would help address this gap.

To address this gap, we implemented a reward-allocation task in pairs of rhesus macaques while at the same time recording from single neurons in three critical nodes in the decision-making network, namely the anterior cingulate gyrus (ACCg), the anterior cingulate sulcus (ACCs), and the orbitofrontal cortex (OFC). Our study capitalized on monkeys' willingness to engage with a social partner via an interposed computer system while at the same time controlling the sensory and reward environment. We specifically matched choices for the reward outcomes directly received by the actor monkey and controlled for potential secondary acoustic reinforcement effects associated with delivering juice to the recipient monkey (see below). Under these conditions, we found regional biases in the encoding of social decision outcomes with respect to self and another individual. Within this network of received (OFC) and foregone (ACCs) reward signals, ACCg emerges as a key nexus for the computation of shared experience and social reward.

Summary of behavior in the reward-allocation task

On one half of trials, termed *choice trials*, actor monkeys chose between visual stimuli that led to juice delivered either to themselves (*self* reward), to the recipient monkey (*other* reward), or to neither monkey (*neither* reward). Offers appeared in pairs of three types, which defined *Self:Neither* trials, *Self:Other* trials, and *Other:Neither* trials (**Fig. 1a–d**). On the other half, termed *cued trials*, monkeys observed a single cue that indicated *self*, *other* or *neither* rewards would be delivered by the computer, as defined above.

Actors performed the task well (**Fig. 2a**), as indicated by the low mean number of incomplete trials per session (4.6 ± 0.2% [s.e.m.]) (**Online Methods**), even when they had no chance of obtaining juice rewards themselves, which was the case for *Other:Neither* choice trials

and for *other* and *neither* cued trials (7.4 ± 0.3%). Actors also made significantly fewer errors when they made active decisions (choice trials) then when there was no choice (cued trials), when there was no reward at stake for themselves ($P < 0.0001$, Welch two sample *t*-test). These findings suggest monkeys find it rewarding to actively choose what to do, and can be motivated to work without direct reinforcement.

Reaction times often serve as a proxy for motivation in incentivized tasks[29–33]. Reaction times for making different choices demonstrate that actors discriminated the reward types and had orderly preferences amongst them[29,33]. Actors were fastest to choose *self* rewards, followed by *other* rewards and *neither* rewards (**Fig. 2b**). *Self* vs. *other* reaction times differed by a mean of 39 ms ($P < 0.0001$; Welch two-sample *t*-test); *other* vs. *neither* differed by a mean of 20 ms ($P < 0.0001$). The ordered reaction times by monkeys making choices in the reward allocation task suggest that rewarding self is more reinforcing than rewarding the recipient, which is in turn more reinforcing than rewarding no one[33].

Finally, actors shifted gaze to the recipients more frequently following juice delivery to them compared to juice delivery to themselves or to neither monkey, consistent with greater interest in the actions of the other monkey when he was rewarded (**Supplementary Figure 1**). Taken together, these observations support the conclusion that actors were acutely aware of the difference between *self*, *other*, and *neither* reward outcomes[33].

We quantified decision preferences by calculating a contrast ratio based on actors' choices (**Online Methods** Eq. 1). Consistent with our previous reports[33,34], actors preferred *self* rewards over *other* or *neither* rewards, but preferred *other* over *neither* rewards (**Fig. 2c**). On *Self:Neither* and *Self:Other* trials, actors almost always chose to reward *self* (**Fig. 2c**) (preference index [mean ± s.e.m.]: *Self:Neither*, –0.99 ± 0.00; *Self:Other*, –0.99 ± 0.00; significantly

different from zero: both $P < 0.0001$, one sample *t*-test). By contrast, on *Other:Neither* trials, actors preferred to allocate rewards to the recipient monkey (**Fig. 2c**) (0.17 ± 0.01; $P < 0.0001$, one sample *t*-test). We observed similar choice preferences for each actor individually (**Supplementary Figure 2**).

We previously reported that the preference to allocate reward to the other monkey is enhanced by greater familiarity between the two animals, and is abolished if the recipient is replaced with a juice collection bottle[33]. We also reported that reward withholding is reduced when actor monkeys are dominant toward recipients, and the variability and the degree of preferences often depend on the identity of the recipients[33]. Furthermore, we reported that actor monkeys prefer to deliver juice to themselves compared to both themselves and the recipient simultaneously, perhaps reflecting the competitive nature of simultaneously drinking juice—a resource controlled outside of experimental sessions in order to motivate performance and often monopolized by dominant monkeys living in pairs with subordinate monkeys in their home cages[33] (MLP, personal observation). Finally, exogenously increasing oxytocin levels in the central nervous system amplifies actors' preference to allocate reward to the other monkey over no one[34]. Taken together, these patterns of behavior endorse the fundamentally social nature of the reward-allocation task.

We also found that preferences scaled with the magnitude of juice on offer. With larger amounts of juice at stake, actors became more motivated to receive (*Self:Neither* & *Self:Other*, slope significantly different from zero: both $P < 0.001$, type II regression) and also to allocate rewards to the other monkey over no one (*Other:Neither*, $P < 0.05$) (**Fig. 2d**). These findings suggest that both direct and vicarious reinforcement processes that motivate social decisions are magnified by reward magnitude[25–27].

Differential encoding of social decision outcomes

We recorded the activity of single neurons in ACCg (n = 81), ACCs (n = 101), and OFC (n = 85) from two actor monkeys (**Fig. 3a**) during the reward-allocation task. We describe neuronal responses from typical single neurons and the populations below for each region. We analyzed the data for both a choice/cue epoch and a reward epoch (**Online Methods**). **Supplementary Figure 3** shows population data for the individual monkeys. For brevity, we focus on the reward epoch; data for the choice/cue epoch are found in **Supplementary Figure 4**, as well as in **Figures 3** and **4**. Overall, we found remarkable resemblances in activity and functional classes (see below) across the choice and reward epochs.

***ACCg*:** ACCg contained neurons selective for allocating rewards to another individual, receiving rewards, or both. One class of ACCg neuron (**Fig. 3b**) preferentially responded when actors chose to allocate reward to recipients. On choice trials, this example neuron discharged more strongly when the actor chose *other* rewards (7.12 ± 0.66 [mean and s.e.m.] spikes/s [sp/s]) compared to *self* rewards on either *Self:Neither* or *Self:Other* trials (4.95 ± 0.36, 4.93 ± 0.45 sp/s, respectively) (both $P < 0.01$, Welch two sample t-test), and also preferred *other* rewards over *neither* rewards (4.44 ± 0.79 sp/s, $P < 0.05$). This neuron did not differentiate *self* from *neither* rewards ($P = 0.97$, Welch two sample t-test). On cued trials, this neuron only weakly preferred *other* over *self* or *neither* rewards (both $P = 0.08$, Welch two sample t-test) (**Fig. 3b**).

By contrast, another class of ACCg neuron (example neuron in **Fig. 3c**) responded selectively for choosing *self* rewards. The example neuron in **Figure 3c** neuron discharged more when the actor chose to reward himself on *Self:Neither* and *Self:Other* trials (4.77 ± 0.38, 5.70 ±

0.41 sp/s, respectively) compared to choosing *other* and *neither* rewards (2.02 ± 0.32, 1.60 ± 0.39 sp/s, respectively) (all $P < 0.0001$, Welch two sample *t*-test). Moreover, it showed stronger responses when the actor received rewards in *Self:Other* than *Self:Neither* context, but this effect did not reach statistical significance ($P = 0.10$, Welch two sample *t*-test). On cued trials, this neuron preferred *self* over *other* or *neither* rewards (both $P < 0.0001$, Welch two sample *t*-test). For both choice and cued trials, the response did not differentiate *other* and *neither* rewards (both $P > 0.23$, Welch two sample *t*-test).

Finally, a third class of ACCg neuron (example neuron in **Fig. 3d**) responded equivalently to both received rewards (*Self:Neither*, 15.28 ± 0.70, *Self:Other*, 16.47 ± 0.81 sp/s) and allocated rewards to other (15.81 ± 1.16 sp/s) (both $P > 0.64$, Welch two sample *t*-test), but responded significantly less to *neither* rewards (10.17 ± 1.23 sp/s; *other* vs. *neither* and *self* vs. *neither*: both $P < 0.005$). Similarly, on cued trials, this neuron preferred *other* over *neither* rewards ($P < 0.05$, Welch two sample *t*-test), but did not differentiate between *self* and *other* rewards ($P = 0.27$).

Importantly, the fact that the solenoid valves controlling juice delivery (including one for *neither* rewards that only produced clicks) were placed outside the experimental room, as well as the white noise played inside the room, during sessions rules out a simple explanation that *other*-reward specific (**Fig. 3b**) and shared *self*/*other* reward responses (**Fig. 3d**) were merely sensory responses to the sounds of the reward-delivery mechanism.

To contrast population coding of decision and reward information in various conditions, we computed a normalized activity bias between each pair of outcomes, expressed as a proportional modulation in mean firing rates normalized by baseline firing rate. In the ACCg population, the mean normalized activity bias for *other* over *neither* rewards (*other* vs. *neither*)

was 0.21 ± 0.10 (s.e.m.), i.e., a 21% difference, which was significant ($P < 0.05$, paired *t*-test) (**Fig. 3e**, **5a**). Similarly, the bias for *self* (from *Self:Other*) over *neither* rewards was 0.20 ± 0.12 ($P = 0.09$, paired t-test). Notably, the population showed equivalent responses for *self* rewards (*Self:Other*) and *other* rewards (0.01 ± 0.12, $P = 0.96$, paired *t*-test). On the other hand, it showed a significant bias for *self* rewards when the actors were presented with a choice between rewarding themselves and recipients compared to when the actors were presented with a choice between rewarding themselves and no one (*Self:Other* vs. *Self:Neither*, by 0.17 ± 0.08, $P < 0.05$, paired *t*-test), suggesting that ACCg is particularly sensitive to a reward context involving an option to reward another individual. Thus, the ACCg population showed an equivalent preference for *other* and *self* rewards, and preferred both over *neither* rewards.

On cued trials, however, a strikingly different pattern emerged. The population responded strongly to *self* rewards but barely responded to *other* rewards (0.59 ± 0.32, $P = 0.07$, paired *t*-test) (**Fig. 3e**). Furthermore, the population now responded no differently to *other* and *neither* rewards (0.22 ± 0.14, $P = 0.14$, paired *t*-test).

Taken together, these results indicate that ACCg, as a population, encodes both giving and receiving rewards. At the population level, neuronal activity selective for allocating rewards to another individual is specific to active decisions (upper vs. lower: **Fig. 3e**), similar to what has been reported by fMRI of human ventral striatum during voluntary versus forced charitable donations[25]. The confluence of neurons selectively responsive to *self*, *other*, and both (*self* and *other*) rewards in ACCg suggests this area contains the information necessary to mediate the vicarious reinforcement processes that appear to motivate actors to give to recipients.

***ACCs*: Fig. 4a** shows a typical ACCs neuron that fired more strongly preceding *other* and *neither* rewards than *self* rewards. On choice trials, this neuron discharged more strongly

when the actor chose not to reward himself (*other* rewards, 19.64 ± 2.15; *neither* rewards, 18.19 ± 2.03 sp/s) compared to when he chose to reward himself directly (*Self:Neither*, 10.31 ± 0.86; *Self:Other*, 9.79 ± 0.81 sp/s) (all $P < 0.001$, Welch two sample *t*-test). This neuron responded equivalently to *self* rewards in *Self:Other* and *Self:Neither* contexts ($P = 0.66$, Welch two sample *t*-test), and also responded equivalently to *other* and *neither* rewards ($P = 0.62$), consistent with encoding "foregone" rewards. On cued trials, this neuron responded equivalently to *other* and *neither* rewards ($P = 0.39$, Welch two sample *t*-test), but less to *self* rewards (both $P < 0.005$), resembling the responses to active decisions.

Likewise, the ACCs population showed a strong and equivalent response bias for foregone rewards (*self* vs. *other*, activity bias=0.31 ± 0.07; *self* vs. *neither*, activity bias=0.25 ± 0.08, both $P < 0.005$, paired *t*-test) (**Fig. 4c**, **5b**). The population did not differentiate *other* from *neither* rewards (0.06 ± 0.06, $P = 0.31$, paired *t*-test). Unlike ACCg, the population did not respond differentially to *Self:Other* and *Self:Neither* contexts (differed by 0.003 ± 0.02, $P = 0.90$, paired t-test). We found similar patterns on cued trials – responses to *self* rewards were substantially reduced compared to *other* rewards (0.19 ± 0.09, $P < 0.05$, paired *t*-test) and *neither* rewards (0.18 ± 0.10, $P < 0.08$) (**Fig. 4c**). These results indicate that, during social interactions, ACCs neurons predominantly signal foregone rewards.

***OFC*: Fig. 4b** shows a typical OFC neuron that preferentially encoded juice rewards received by the actor. On choice trials, this neuron discharged significantly more for *self* rewards than for the alternatives on both *Self:Neither* and *Self:Other* trials. Activity for *self* rewards did not differ between the two *self* reward contexts (7.00 ± 0.47, 7.03 ± 0.46 sp/s, respectively, $P = 0.97$, Welch two sample *t*-test), but it exceeded the cell's activity for *other* and *neither* rewards (3.06 ± 0.40, 1.85 ± 0.42 sp/s, respectively; both $P < 0.0001$). On cued trials, this neuron

responded most strongly to *self* rewards compared to both *other* and *neither* rewards (both $P < 0.0001$, Welch two sample *t*-test), but it did not respond differently between *other* and *neither* rewards ($P = 0.25$) (**Fig. 4b**).

The OFC population predominantly encoded *self* rewards compared to *other* and *neither* rewards. The bias for *self* over *other* rewards was 30% (0.30 ± 0.09, $P < 0.005$, paired *t*-test). For *self* versus *neither* rewards, the bias was also significant (0.17 ± 0.08, $P < 0.05$, paired *t*-test) (**Fig. 4d**, **5c**). Population activity for *other* and *neither* rewards did not differ (0.08 ± 0.06, $P = 0.20$, paired *t*-test) (**Fig. 4d**, **5c**). Unlike ACCg, the population did not respond differentially to *Self:Other* and *Self:Neither* contexts (differed by 0.06 ± 0.07, $P = 0.39$, paired t-test). On cued trials, the *self* reward bias was not present compared to *other* rewards (0.19 ± 0.16, $P = 0.24$, paired *t*-test) and only weakly present over *neither* rewards (0.26 ± 0.15, $P < 0.08$). On cued trials, the population did not distinguish *other* rewards from *neither* rewards ($P = 0.33$, paired *t*-test) (**Fig. 4d**). These results indicate that OFC neurons predominantly encode rewards received by the actors, and this information was encoded more faithfully during active decision-making.

Neuronal reference frames for social decisions

Neuroimaging and scalp-recording studies in humans can only study neuronal activity at an aggregate level. Our single-unit recording data thus provide a unique opportunity to quantify the frame of reference in which individual neurons within ACCg, ACCs, and OFC encode social decisions. To do this, we classified cells from each area based on an analysis of variance (ANOVA) of neuronal activity of individual neurons with reward outcome (*self*, *other*, or *neither*), trial type (choice or cued), and reward magnitude (small, medium, or large) as factors (**Online Methods**). Reward epoch responses differed significantly for a large number of neurons

from all areas in a manner that depended on reward outcome (ACCg: 57%, ACCs: 72%, OFC: 57%), trial type (ACCg: 36%, ACCs: 52%, OFC: 45%) and reward volume (ACCg: 12%, ACCs: 25%, OFC: 24%) (**Supplementary Table 1**). Furthermore, we observed remarkable resemblances in reward outcome coding across the choice/cue and reward epochs (**Supplementary Figure 4**).

Based on the statistical significance of the ANOVA during the choice/cue and reward epochs, we classified individual neurons as self-referenced (i.e., modulation referenced to *self* rewards, preferring either self or foregone rewards), other-referenced (i.e., modulation referenced to *other* rewards), both-referenced (i.e., modulation referenced to both *self* and *other* rewards, but not *neither* rewards), or unclassified (**Online Methods**). Here we consider the proportion of different cell types among the classified neurons based on this scheme. In OFC, 80% (n = 36/45) were self-referenced, whereas only 9% (4/45) were other-referenced and 11% (5/45) were both-referenced (both $P < 0.0001$, χ^2 test). In ACCs, 72% (51/71) were self-referenced, whereas only 14% (10/71) were other-referenced and 14% (10/71) were both-referenced (both $P < 0.0001$, χ^2 test) (**Fig. 5d**). In contrast, ACCg contained similar proportions that were self-referenced (38%, 12/32), other-referenced (31%, 10/32), and both-referenced (31%, 10/32) ($P > 0.79$, χ^2 test). Critically, ACCg contained a significantly higher proportion of neurons (>60%) that were sensitive to the reward outcome of the recipient monkey (i.e., other-referenced and both-referenced) compared to either OFC or ACCs (both $P < 0.005$, χ^2 test) (**Fig. 5d**). ACCg also contained a significantly smaller proportion of self-referenced neurons than either OFC or ACCs (both $P < 0.005$, χ^2 test). Finally, we found similar results when we repeated the analysis by including trial-by-trial choice reaction times as covariates (**Supplementary Figure 5**).

To test whether different neuronal frames of reference (self-, other-, and both-referenced) were anatomically segregated, we used principal component analysis on recording coordinates to identify the major axis with the largest dispersion within three-dimensional space. We then projected neurons to that axis to test differential distributions in individual monkeys separately. **Figure 6** shows reconstructed recording locations for each reference frame class for each area. We did not observe any systematic anatomical clustering amongst different frames of reference: self-, other-, and both-referenced neurons within ACCg, ACCs, and OFC were intermingled (all $P > 0.56$, Wilcoxon rank sum test).

Next we examined whether differential encoding of self, other, and neither rewards was also present prior to making a decision. We found very little evidence for systematic signals early in the trial just, after target onset (from 50ms to 250ms from target onset). In ACCg, only 0, 3, and 1 cells were classified into self-, other-, and both-referenced classes with only 12% of neurons showing significant effect of reward type. In ACCs, only 1, 2, and 3 cells belonged to each category, with only 22% of the neurons with significant reward type effects. Similarly, in OFC, only 2, 2, and 4 cells belonged to each category, with only 28% of the neurons with significant reward type effects. Thus, in our reward allocation task, signals in ACCg, ACCs, and OFC appear to emerge around the time of choice and reward delivery.

Finally, we examined whether session-to-session variation in prosocial tendencies on *Other:Neither* trials (**Fig. 2c**) could be explained by variability in the responses of ACCg neurons—the population most sensitive to other's rewards. We split recording sessions based on actors' choices on *Other:Neither* into two categories: more prosocial (higher *other* over *neither* choices relative to the median preference index) and less prosocial (lower *other* over *neither* choices relative to the median preference index). Actors tended to be more prosocial on

recording sessions when other-referenced and both-referenced ACCg neurons showed less variability in spiking during the reward epoch ($P < 0.05$, bootstrap test) (**Fig. 7a**). By contrast, we found that self-referenced ACCg neurons generated more variable responses during the reward epoch when actors were more prosocial ($P < 0.05$, bootstrap test). ACCs neurons did not show any systematic relationship between response variance and behavior ($P = 0.47$, bootstrap test; **Fig. 7b**). Notably, OFC neurons showed a similar pattern to self-referenced ACCg neurons ($P < 0.005$, bootstrap test; **Fig. 7c**). These findings reveal suggest a strong link between prosocial behavior and the fidelity of social reward signals carried by those neurons that incorporate the experience of others into their responses. This could be due to enhanced attention to the recipient or other processes known to influence signal to noise in cortical neurons.

Discussion

Our findings strongly endorse the hypothesis that distinct frontal regions contribute uniquely to social decisions by differentially processing decision outcomes with respect to actors (*self*) and their partners (*other*). The finding that OFC neurons selectively encode self reward is consistent with previous studies implicating this area in representing the subjective value of rewards[12,13], but extend those findings by demonstrating that such value signals are encoded egocentrically. Encoding of foregone rewards by ACCs neurons, on the other hand, is consistent with previous data implicating this area in error monitoring and behavioral adjustment[35–37]. For example, foregone reward signaling by ACCs might be used to learn from observation, rather than direct experience, and adjust ongoing behavior during social interactions. Furthermore, mirroring of *self* and *other* rewards by ACCg neurons is consistent with previous studies linking this area to specifically social functions like shared experience and empathy[38].

Our findings echo those of a previous study examining the effects of lesions in these same brain regions (**Online Methods**), which demonstrated that ACCg, but not OFC or ACCs, contributes causally to the use of visual social information to guide behavior[9]. Specifically, ACCg lesions completely abolished typical hesitation to retrieve food when confronted with social stimuli[9]. Our findings also agree with previous findings that lesions in ACCs impair the use of reward history to guide decisions adaptively[10]. Differences between ACCs and ACCg reported here support and extend the finding that learning based on experience is mediated by ACCs, whereas learning from feedback from another individual is mediated by ACCg[8]. Specifically, in a learning task in which human subjects monitored their history of correct responses as well as the advice given to them by a confederate, BOLD activation in ACCs tracked reward learning rate, whereas BOLD activation in ACCg tracked social learning rate based on advice from the confederate[8]. In our study, we propose that ACCs tracked foregone rewards relative to self, whereas ACCg tracked reward outcomes of another individual in a more complex manner.

Intriguingly, the ACCg population also responded more strongly when monkeys chose *self* reward when the alternative was allocating reward to the other monkey compared to the response when monkeys chose *self* reward when the alternative was rewarding no one. In contrast, neither the OFC neuronal population response nor the ACCs neuronal population response was sensitive to social context when monkeys rewarded themselves. Sensitivity to social context in ACCg endorses a specialized role for this area in computing social decisions – even when one acts selfishly.

It is worthwhile to note that a small number of ACCs and OFC neurons, though much less in proportion compared to ACCg (**Fig. 5d**, **Supplementary Table 1**, **Supplementary**

Figure 5), were classified as either other- or both-referenced. This observation supports the idea that a small number of ACCs and OFC neurons do carry information about rewards allocated to another individual. What is striking here is that the majority of OFC and ACCs neurons (80% and 72%, respectively) do not carry such other-regarding information (other- or both-referenced), whereas the majority of ACCg neurons do so (62%). This endorses a fundamentally social role for neurons in ACCg.

A prior study showed that OFC neurons modulate their activity when a monkey receives juice reward together with another individual[39], suggesting that value signals in OFC are sensitive to social context. In that study, OFC neurons responded differentially as a function of whether the subject monkey received juice rewards alone or together with another monkey[39]. Our current study builds upon and extends those findings in three important ways. First, we used a free-choice task that allowed us to infer the subjective value of rewards delivered to self, other, and no one. Remarkably, even in a social context OFC neurons were selective for *self* reward, the most preferred outcome. Second, we compared the responses of OFC neurons to responses of neurons in ACCg and ACCs recorded in identical task conditions, allowing us to demonstrate regional differences in the encoding of social reward information in primate frontal cortex. Third, when we compared responses of ACCg neurons on free-choice and cued trials we found that responses to rewards delivered to the recipient monkey were largely absent when actors passively observed the event rather than actively choosing it. Taken together, these extensions demonstrate that social context can impact the encoding of reward information in all three areas: OFC appears to be implicated with the evaluation of personally experienced rewards, ACCs evaluates reward information that is not directly experienced, and ACCg multiplexes information

about the direct experience of reward and vicarious reinforcement experienced by allocating reward to another individual.

It is noteworthy that ACCs neurons showed much less modulation by actors' received reward outcomes compared to OFC neurons. This is striking since ACCs neurons often show substantial modulation to received reward in nonsocial settings[11]. ACCg, on the other hand, contains neurons that compute reward signals in both *other* and *self* frames of reference. Together, our findings suggest that, as in sensory and motor systems[40], identifying the frames of reference in which reward outcomes are encoded may be important for understanding the neural mechanisms underlying social decision-making[8].

Accumulating evidence endorses a special role for the medial-frontal cortex in representing information about another individual[8,41–44]. For instance, perceived similarity while observing others is correlated with hemodynamic response in the subgenual ACC[44]. Further, a group of neurons in the primate medial-frontal cortex selectively responds to observing actions performed by other individuals[41]. Such other-referenced signals, however, are not limited to the medial wall of the frontal cortex. Neurons in the dorsolateral prefrontal cortex (DLPFC) track the behavior of a computer opponent in an interactive game[45], and BOLD responses in DLPFC and ventromedial prefrontal cortex (vmPFC) during observational learning track observed action and observed reward prediction errors, respectively[46]. Furthermore, BOLD activity in anterior frontal areas tracks preferences to donate to charity[24]. Brain networks involved in mentalizing[47], vicarious pain perception[48] and empathy[49] thus seem to be critical for mediating social interactions, suggesting that other-regarding cognition is orchestrated by a distributed network of frontal cortical areas.

Social and emotional behaviors are highly idiosyncratic among individuals. Understanding the neural mechanisms that drive such individual differences remains one of the most pressing issues in neuroscience. We hypothesize that the differential activation of neurons in ACCg, ACCs, and OFC contribute to individual and, perhaps species, differences in social function.

Supplementary Information

Supplementary Information includes Supplementary Table and Figures.

Acknowledgements

This work was supported by T32 Postdoctoral Training Grant on Fundamental and Translational Neuroscience (S.W.C.C.; 2T32NS051156–06), Ruth K. Broad Biomedical Foundation Postdoctoral Grant (S.W.C.C.), Canadian Institutes of Health Research Doctoral research award (J.F.G.; 84765), National Institute of Mental Health (M.L.P., S.W.C.C.; MH095894), and Department of Defense (M.L.P., S.W.C.C.; W81XWH-11-1-0584). We thank Jennifer M. Groh, John M. Pearson, David L. Barack, R. Becket Ebitz, Ethan S. Bromberg-Martin, Karli K. Watson, and Benjamin Y. Hayden for helpful discussions. We are very grateful to Steven P. Wise and Camillo Padoa-Schioppa for insightful discussions and comments on the earlier versions of the manuscript. We also thank Monica L. Carlson for general technical assistance.

Author Contributions

S.W.C.C and M.L.P. designed the research and wrote the paper. S.W.C.C. and J.F.G. performed the research and S.W.C.C analyzed the data.

Author Information

The authors declare that they have no competing financial interests. Correspondence and requests for materials should be addressed to S.W.C.C. (e-mail: steve.chang@duke.edu).

Figure Legends

Figure 1 Reward-allocation task. (**a**) Experimental setup for an actor and a recipient monkey. (**b**) Stimulus-reward outcome mappings for reward delivered to actor (*self*), recipient (*other*), or no one (*neither*), shown separately for each actor. (**c**) Magnitude cue used to indicate juice amount at stake for each trial (see **d**). Position of the horizontal bisecting line specified the percentage of maximum reward possible. (**d**) Task structure (see **Online Methods**). Top fork, cued trials; bottom fork, choice trials. Dashed gray lines show the angle of the actor's gaze, converging on the fixation point. Eye cartoons indicate times when the actor could look around. RT, reaction time. MT, movement time. ITI, inter-trial interval.

Figure 2 Behavior in the reward-allocation task. (**a**) Proportions of incomplete trials (mean ± s.e.m.) (see **Online Methods**) during the reward-allocation task. (**b**) Choice reaction times (ms) from trials in which rewards were chosen for *self*, *other*, or *neither* (mean of session medians ± s.e.m). (**c**) Choice preferences (preference index, mean ± s.e.m.) as a function of reward outcome contrasts. Data points next to each bar show means for individual sessions. The degree of preference axis on the right shows the range of preference indices in ratio terms. (**d**) Choice preferences (mean ± s.e.m.) as a function of reward magnitude on 219 single-unit sessions collected with the magnitude cue.

Figure 3 Single neurons and population responses from ACCg. (**a**) Structural magnetic resonance image from actor MO, with example electrode paths for ACCg, ACCs and OFC. (Also see **Fig. 6**.) (**b**) Mean responses (peri-stimulus time histograms [PSTHs]) and spike rasters for an

other-reward preferring ACCg neuron, on choice trials (upper, solid traces) and cued trials (lower, dashed traces). Data are aligned to choice/cue offset (left) and reward onset (right) for each reward outcome. Bar histograms on right show mean ± s.e.m. activity from the two epochs (grey regions). Color codes for PSTH traces and histograms shown below. (**c**) PSTHs and spike rasters for a *self*-reward preferring ACCg neuron. (**d**) PSTHs and spike rasters for a shared *self* and *other* reward preferring ACCg neuron. (**e**) Normalized choice/cue epoch and reward epoch responses for 81 ACCg neurons. **c–e**, same format as in **b**. In all bar histogram insets, the horizontal lines above different conditions indicate significance differences (solid, $P < 0.05$ by paired *t*-test; dashed, $P < 0.05$ by bootstrap test).

Figure 4 Single neurons and population responses from ACCs and OFC. (**a**) PSTHs and spike rasters for a single ACCs neuron preferring forgone rewards. Data are aligned to choice/cue offset (left) and reward onset (right) for each reward outcome. Bar histograms on right show mean ± s.e.m. activity from the two epochs (grey regions). (**b**) PSTHs and spike rasters for a single OFC neuron preferring *self* reward. (**c**) Normalized reward epoch responses of 101 ACCs neurons. (**d**) Normalized choice/cue epoch and reward epoch responses of 85 OFC neurons. All panels, same format as in **Fig. 3**.

Figure 5 Population biases for *self*, *other*, and *neither* rewards. Scatter plots show mean normalized reward epoch responses (proportion of modulation relative to baseline) of individual neurons (from left to right) between *self* (*Self:Other*) and *other* rewards, between *other* and *neither* rewards, between *self* rewards from *Self:Neither* and *Self:Other* contexts, and between *self* (*Self:Neither*) and *neither* rewards, for ACCg (**a**), ACCs (**b**), and OFC (**c**) populations.

Regression lines (type II) are shown in red (the circled data points are excluded from the regression). Unity lines are shown in black. The example neurons from **Fig. 3,4** are indicated on the scatter plots. (**d**) Proportion of neurons (out of significantly classified neurons) from OFC, ACCs, and ACCg using self-referenced, other-referenced, and both-referenced frames to represent reward outcomes. Inset shows color codes used in the bar graph. Bars indicate significant differences in proportions ($P < 0.05$, χ^2 test).

Figure 6 Anatomical projections of recorded locations of all ACCg, ACCs, and OFC cells. Recording sites were transformed from chamber coordinates into interaural coordinates. The interaural coordinates of individual cells from both monkeys were then projected onto standard stereotaxic maps of rhesus monkeys[50], with a 2 mm interaural spacing in the anterior-posterior dimension. Cells are shown on coronal slices and color-coded for the types of frames of reference used, as specified in **Supplementary Table 1** (see box). The lateral view of the brain (inset) shows the locations of the coronal sections. cgs, cingulate sulcus; ps, principal sulcus; morb, medial orbitofrontal sulcus; lorb, lateral orbitofrontal sulcus; ros, rostral sulcus; Cd, caudate.

Figure 7 Prosocial behavior and the fidelity of neuronal responses on *Other:Neither* trials. (**a**) ACCg; (**b**) ACCs; (**c**) OFC. Coefficients of variation in firing rate (CV; **Online Methods**) during the reward epoch on *other* reward trials are plotted as a function of whether actors were more or less prosocial on *Other:Neither* trials based on median split (higher: preference index greater than median; lower: preference index less than median). Asterisks indicate $P < 0.05$ by bootstrap test.

Online Methods

General and behavioral procedures

All procedures were approved by the Duke University Institutional Animal Care and Use Committee, and were conducted in compliance with the Public Health Service's Guide for the Care and Use of Laboratory Animals.

Two actor (MY and MO) and five recipient monkeys (*Macaca mulatta*) participated. For all monkeys, a sterile surgery was performed to implant a head-restraint prosthesis (Crist Instruments) using standard techniques[11]. Six weeks after surgery, monkeys were trained on a standard, center-out, oculomotor task for liquid rewards. Actor monkeys were then trained on the reward-allocation task (**Fig. 1a–d**) in the presence of a recipient. Subsequently, a second surgery was performed on actors to implant a recording chamber (Crist) providing access to both the sulcal and gyral regions of the anterior cingulate cortex (ACCs and ACCg, respectively) and the orbitofrontal cortex (OFC). All surgeries were performed under isoflourane anesthesia (1–3%), and the recording chambers were regularly cleaned, treated with antibiotics and sealed with sterile caps.

Horizontal and vertical eye positions were sampled at 1000Hz using an infrared eye monitor camera system (SR Research Eyelink). Stimuli were controlled by PsychToolBox and Matlab (Mathworks). Actors and recipients sat in primate chairs (Crist), 100cm from one another at a 45-degree angle (**Fig. 1a**). Actors (both males) and recipients (four males, one female) were unrelated and not cagemates. Different pairs were selected depending on the availability of recipient monkeys. Actors were housed in a colony with 12 other male rhesus macaques, some of which were pair-housed. All the male monkeys resided in this colony room, and the one female monkey resided in the adjacent colony room with other females. Out of the total seven actor-

recipient pairs tested in the current study, the actor monkey was dominant over the recipient in six cases. Furthermore, three pairs could be classified as "more familiar" with one another because their cages faced each other, as defined previously[33]. Based on these relationships, we would expect a mixture of prosocial and competitive preferences based on our prior results showing dominant actors are slightly less competitive than subordinates, but also showing that pairs in which the actor is less familiar with the recipient are slightly less prosocial than when they are more familiar.

In the experimental setup, each monkey had his own monitor which displayed identical visual stimuli. Both the actor and recipient monkeys had their own tube from which juice drops were delivered. In order to prevent monkeys from forming secondary associations of solenoid valve clicks or the sound of the recipient drinking the juice reward with respect to different reward types, the solenoid valves that delivered the juice rewards were placed in another room and white noise was also played in the background. Experimenters were unable to hear solenoids anywhere inside the recording room. Our control of the acoustic environment explicitly rules out a simple explanation that both-referenced reward encoding found in ACCg is a product of such secondary sensory associations. Critically, a separate solenoid (also placed in another room) was designated for *neither* rewards; it only produced clicks but delivered no fluid.

The face region of the recipient, with respect to the gaze angle of the actor (horizontal and vertical eye positions), was determined empirically prior to the experiments. The frequency with which actors looked at recipients was computed from number of gaze shifts to the recipient's face (±8.5° from the center of the face)[33,34]. We used a large window to capture gaze shifts that were brief in duration and large in magnitude and often directed at varying depths (e.g., eyes, mouth) (**Fig. 1a**).

Monkeys performed the task to obtain drops of cherry- or orange-flavored juice. Actors began a trial by shifting gaze (±2.5˚) to a central stimulus (0.5˚ x 0.5˚), and maintained fixation (200ms). For 219 single-unit sessions, the reward magnitude at stake (0.1 – 2.4ml) on each trial was cued by the position of a horizontal bisecting line (200ms), indicating the percentage of the maximum possible volume. There were two kinds of trials, termed *choice trials* and *cued trials*. Following a variable delay (300, 500, 700ms), choice and cued trials were presented at equal probabilities, randomly interleaved. On choice trials, two visual targets (4˚ x 4˚) appeared at two random locations 7˚ eccentric in the opposite hemifield. Actors shifted gaze to one target (±2.5˚) to indicate a choice within the maximum allowed time of 1.5s (from stimulus onset). The pair of stimuli appearing on a given trial was drawn from the set of three stimuli (**Fig. 1b**), pseudorandomly selected. On cued trials, actors maintained fixation (±2.5˚) while a cue (4˚ x 4˚) appeared centrally (500ms). Cues indicating rewards for the actor, recipient or neither monkey occurred with equal frequency, pseudorandomly determined (**Fig. 1b**). Reward onset was followed by a 0–900ms delay, from the time of either making a choice or cue offset. Actors were free to look around during this delay and for one second after reward delivery. Reward delivery was followed by an intertrial interval of 700, 1,000, or 1,300ms. Upon making an error (see below), both monkeys received visual feedback (a white rectangle, 10˚ x 10˚) followed by a 5s time out before the next trial.

Recording procedures

All recordings were made using tungsten electrodes (FHC). Single electrodes were lowered using a hydraulic microdrive system (Kopf Instruments, or FHC). Single-unit

waveforms were isolated, and action potentials collected, using a 16-channel recording system (Plexon).

In order to guide the placement of recording tracks and localize recording sites, we acquired structural magnetic resonance images (MRI) (3T, 1 mm slices) of each actor's brain. Detailed localizations were made using Osirix-viewer. In addition to MRI-guidance, we confirmed that electrodes were in ACCg, ACCs, or OFC by listening to grey-matter and white-matter associated sounds while lowering the electrodes. ACCg neurons were recorded from Brodmann areas 24a, 24b and 32; ACCs neurons (dorsal and ventral banks) were recorded from 24c and 24c'; OFC neurons were recorded from 13m and 11 (based on standard anatomical references[51,52]) (**Fig. 3a** and **Fig. 6**).

Single-unit recordings were made from two actor monkeys while each was engaged in a reward-allocation task with a recipient monkey in 267 sessions. A total of 81 ACCg neurons (MY: 45, MO: 36), 101 ACCs neurons (MY: 39, MO: 62), and 85 OFC neurons (MY: 46, MO: 39) were included in the study. Neurons were selected for recording based solely on the quality of isolation. For a small subset of the data (18% of the total) (ACCg: 0%; ACCs: 25%; OFC: 27%), data were collected in a task with a fixed reward size (typically 1.0ml per successful trial) (identical to **Fig. 1d** except without the magnitude cue). For the majority of the cells (82% [n = 219] of the total), data were either collected in a task with the magnitude cue (**Fig. 1d**) (ACCg: 100% [n = 81]; ACCs: 60% [n = 61]; OFC: 42% [n = 36]), or both with and without the magnitude cue (i.e., two or more consecutive blocks per cell) (ACCg: 0%; ACCs: 15% [n = 15]; OFC: 31% [n = 26]). We combined the two types of data in our analyses unless otherwise specified.

Data from each cell consisted of firing rates during 440 ± 13 (±217) (median ± s.e.m. (±s.d.)) trials. A trial was considered "incomplete" if the monkey failed to choose a target on choice trials (choice-avoidance error) or to maintain fixation after cue onset on cued trials (forced-choice avoidance error). Such trials were not included in the neural analysis. The monkeys performed the task well, as evidenced by a high percentage of correct trials even on trials in which they did not receive juice reinforcement (**Fig. 2a**).

Data analysis

Choice preference indices were constructed as contrast ratios (Eq. 1)[33,34].

$$Preference\ Index = \frac{R_A - R_B}{R_A + R_B}. \quad (1)$$

R_A and R_B were the frequency of making particular choices. For *Self:Other* trials, R_A and R_B were number of choices to reward *other* and *self*, respectively. For *Other:Neither* trials, R_A and R_B were number of choices to reward *other* and *neither*, respectively. Finally, for *Self:Neither* trials, R_A and R_B were number of choices to reward *neither* and *self*, respectively. Indices therefore ranged from –1 to 1, with 1 corresponding to always choosing to allocate reward to other on *Other:Neither* trials and *Self:Other* trials, and always choosing not to reward self on *Self:Neither* trials. An index of –1 corresponds to the opposite, generally stated as choosing not to allocate reward to the other monkey or choosing to reward oneself. Values of 0 indicated indifference. For constructing neuronal preferences, we simply substituted the choice frequency with neuronal firing rates associated with making specific decisions. Response times, the time from the onset of choices to movement onset, were computed using a 20°/sec velocity threshold criterion[33,34].

Spike rates were computed during the reward epoch (from 50 to 600ms from reward onset) as well as the choice/cue epoch (from –100 to 300ms from making a choice or cue offset).

For the population analyses, we normalized reward firing rates to the average baseline rates for each reward outcome (300ms interval prior to fixation onset). Using marginally different time windows and different normalization methods all resulted in similar conclusions. Coefficients of variation (CV) were calculated for each neuron based on the standard deviation (σ) and mean (μ) using the spike rates (sp/s) from the reward epoch (Eq. 2):

$$CV = \frac{\sigma}{\mu} \quad (2)$$

In OFC and ACCs populations, the two *self* rewards (i.e., *self* rewards chosen from *Self:Neither* and *Self:Other* trials) were largely indifferent (see **Fig. 4, 5b, 5c** and **Results**) and thus we combined them by taking means for the CV analysis. In contrast, the population of ACCg neurons responded more strongly to *self* rewards obtained from a social context (*Self:Other*) compared to when there was no reward stake for the other monkey (*Self:Neither*) and thus we consider the two *self* rewards separately in ACCg (see **Fig. 3, 5a** and **Results**).

Analysis of variance (ANOVA) was used to classify the reward response selectivity of individual neurons from each area and performed per individual cells. Two-factor ANOVA was used to classify the selectivity of reward outcome (*self*, *other*, or *neither*) and trial type (choice or cued) for all neurons. Three-factor ANOVA was used to classify the selectivity of reward volume (binned into small, medium, large) for the 82% of cells from all areas that were collected in the task with a magnitude cue. Statistical significance for each reward type was computed by Tukey HSD test. Finally, we excluded three OFC cells when our analyses involved using the data from *neither* rewards because these cells were recorded on very rare sessions in which the monkeys either never chose the *neither* reward option or did so fewer than four times. Across all analyses, using slightly different epoch durations for neuronal data analyses led to similar results.

Classification of cell types by significant reward specificity

Based on Tukey HSD tests from the one-way ANOVA on reward outcome (*self*, *other*, or *neither*) for both the choice/cue epoch and reward epoch responses, we classified cells into the following categories: self-referenced, other-referenced, both-referenced, and unclassified. These categories do not imply functional roles but indicate that firing rates were significantly different based on reward outcomes. We refer to a neuron as self-referenced if the responses of the neuron were significantly different ($P < 0.05$) between *self* and *other* rewards as well as between *self* and *neither* rewards, but not different between *other* and *neither* rewards. We refer to a neuron as other-referenced if the responses of the neuron showed significant differences in firing rates between *self* and *other* rewards as well as between *other* and *neither* rewards, but not different between *self* and *neither* rewards. Finally, we refer to a neuron as both-referenced if the responses of the neuron showed significant differences in responses between *self* and *neither* rewards as well as *other* and *neither* rewards, but not different between *self* and *other* rewards. Neurons that did not fall into one of these categories were considered as unclassified. Applying slightly different criteria or differently configured ANOVA did not change the overall proportional trends of these classes.

Reward magnitude analysis

We examined reward magnitude modulation in 219 neurons (i.e., 82% of all neurons collected with the magnitude cue; 81 ACCg, 76 ACCs, and 62 OFC neurons). We performed a linear regression on the activity (sp/s) of individual neurons across unbinned reward sizes. We fit the data using the reward epoch activity separately for *self*, *other*, and *neither* reward outcomes and obtained fitted slopes (i.e., reward magnitude sensitivity in sp/s/ml) for each reward

outcome. For examining the relationship between the reward magnitude sensitivity across actors' received and foregone reward outcomes, we compared the average signed slopes from all received rewards (*self* rewards on choice and cued trials) and all foregone rewards (*other* and *neither* reward on choice and cued trials) in individual neurons.

References

1 Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* 425, 785-791, (2003).
2 Gallese, V., Keysers, C. & Rizzolatti, G. A unifying view of the basis of social cognition. *Trends Cogn Sci* 8, 396-403, (2004).
3 Berger, S. M. Conditioning through vicarious instigation. *Psychological Review* 5, 450-466, (1962).
4 Rilling, J. et al. A neural basis for social cooperation. *Neuron* 35, 395-405, (2002).
5 de Quervain, D. J. et al. The neural basis of altruistic punishment. *Science* 305, 1254-1258, (2004).
6 Baron-Cohen, S., Leslie, A. M. & Frith, U. Does the autistic child have a "theory of mind"? *Cognition* 21, 37-46, (1985).
7 Adolphs, R. Cognitive neuroscience of human social behaviour. *Nature Rev Neurosci* 4, 165-178, (2003).
8 Behrens, T. E., Hunt, L. T. & Rushworth, M. F. The computation of social behavior. *Science* 324, 1160-1164, (2009).
9 Rudebeck, P. H., Buckley, M. J., Walton, M. E. & Rushworth, M. F. A role for the macaque anterior cingulate gyrus in social valuation. *Science* 313, 1310-1312, (2006).
10 Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J. & Rushworth, M. F. Optimal decision making and the anterior cingulate cortex. *Nat Neurosci* 9, 940-947, (2006).
11 Hayden, B. Y., Pearson, J. M. & Platt, M. L. Fictive reward signals in the anterior cingulate cortex. *Science* 324, 948-950, (2009).
12 Tremblay, L. & Schultz, W. Relative reward preference in primate orbitofrontal cortex. *Nature* 398, 704-708, (1999).
13 Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223-226, (2006).
14 Schoenbaum, G., Roesch, M. R., Stalnaker, T. A. & Takahashi, Y. K. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nature Rev Neurosci* 10, 885-892, (2009).
15 Kennerley, S. W., Behrens, T. E. & Wallis, J. D. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci* 14, 1581-1589, (2011).
16 Abe, H. & Lee, D. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* 70, 731-741, (2011).
17 Tsujimoto, S., Genovesio, A. & Wise, S. P. Monkey orbitofrontal cortex encodes response choices near feedback time. J Neurosci 29, 2569-2574, (2009).
18 Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* 275, 1593-1599, (1997).
19 Sato, M. & Hikosaka, O. Role of primate substantia nigra pars reticulata in reward-oriented saccadic eye movement. *J Neurosci* 22, 2363-2373, (2002).
20 Shidara, M., Aigner, T. G. & Richmond, B. J. Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J Neurosci* 18, 2613-2625, (1998).
21 Santos, G. S., Nagasaka, Y., Fujii, N. & Nakahara, H. Encoding of social state information by neuronal activities in the macaque caudate nucleus. *Soc Neurosci* 7, 42-58, (2012).

22 Matsumoto, M. & Hikosaka, O. Representation of negative motivational value in the primate lateral habenula. *Nat Neurosci* 12, 77-84, (2009).
23 Paton, J. J., Belova, M. A., Morrison, S. E. & Salzman, C. D. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* 439, 865-870, (2006).
24 Moll, J. et al. Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc Natl Acad Sci USA* 103, 15623-15628, (2006).
25 Harbaugh, W. T., Mayr, U. & Burghart, D. R. Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316, 1622-1625, (2007).
26 Hare, T. A., Camerer, C. F., Knoepfle, D. T. & Rangel, A. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J Neurosci* 30, 583-590, (2010).
27 Izuma, K., Saito, D. N. & Sadato, N. Processing of the incentive for social approval in the ventral striatum during charitable donation. *J Cogn Neurosci* 22, 621-631, (2010).
28 Kuss, K. et al. A reward prediction error for charitable donations reveals outcome orientation of donators. *Soc Cogn Affect Neurosci*, doi:10.1093, (2011).
29 Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu Rev Neurosci* 30, 535-574, (2007).
30. Sohn, J. W. & Lee D. Effects of reward expectancy on sequential eye movements in monkeys. *Neural Net* 19, 1181-1191 (2006).
31. Bowman, E.M., Aigner, T.G., Richmond, B.J. Neural signals in the monkey ventral striatum related to motivation for juice and cocaine rewards. *J Neurophysol.* 75, 1061-1073 (1996).
32. O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., Dolan, R.J. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304, 452-454 (2004).
33 Chang, S. W., Winecoff, A. A. & Platt, M. L. Vicarious reinforcement in rhesus macaques (Macaca mulatta). *Front Neurosci* 5, 27, (2011).
34 Chang, S. W., Barter, J. W., Ebitz, R. B., Watson, K. K. & Platt, M. L. Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques (Macaca mulatta). *Proc Natl Acad Sci U S A* **109**, 959-964, (2012).
35 Carter, C. S. et al. Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280, 747-749, (1998).
36 Ito, S., Stuphorn, V., Brown, J. W. & Schall, J. D. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* 302, 120-122, (2003).
37 Alexander, W. H. & Brown, J. W. Medial prefrontal cortex as an action-outcome predictor. *Nat Neurosci* 14, 1338-1344, (2011).
38 Amodio, D. M. & Frith, C. D. Meeting of minds: the medial frontal cortex and social cognition. *Nature Rev Neurosci* 7, 268-277, (2006).
39 Azzi, J. C., Sirigu, A. & Duhamel, J. R. Modulation of value representation by social context in the primate orbitofrontal cortex. *Proc Natl Acad Sci U S A* 109, 2126-2131, (2012).
40 Cohen, Y. E. & Andersen, R. A. A common reference frame for movement plans in the posterior parietal cortex. *Nat Rev Neurosci* 3, 553-562, (2002).
41 Yoshida, K., Saito, N., Iriki, A. & Isoda, M. Representation of others' action by neurons in monkey medial frontal cortex. *Curr biol* 21, 249-253, (2011).
42 Saxe, R. Uniquely human social cognition. *Curr Opin Neurobiol* 16, 235-239, (2006).

43 Waytz, A., Zaki, J. & Mitchell, J. P. Response of dorsomedial prefrontal cortex predicts altruistic behavior. *J Neurosci* 32, 7646-7650, (2012).
44 Mobbs, D. et al. A key role for similarity in vicarious reward. *Science* 324, 900, (2009).
45 Seo, H. & Lee, D. Cortical mechanisms for reinforcement learning in competitive games. *Philos Trans R Soc Lond B Biol Sci* 363, 3845-3857, (2008).
46 Burke, C. J., Tobler, P. N., Baddeley, M. & Schultz, W. Neural mechanisms of observational learning. *Proc Natl Acad Sci USA* 107, 14431-14436, (2010).
47 Hampton, A. N., Bossaerts, P. & O'Doherty, J. P. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci USA* 105, 6741-6746, (2008).
48 Jeon, D. *et al.* Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. *Nat Neurosci* **13**, 482-488, (2010).
49 Singer, T. et al. Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157-1162, (2004).
50 Paxinos, G., Huang, X. F. & Toga, A. W. *The rhesus monkey brain in stereotaxic coordinates*. (Academic Press, 2000).
51 Vogt, B. A. & Pandya, D. N. Cingulate cortex of the rhesus monkey: II. Cortical afferents. *J Comp Neurol* 262, 271-289, (1987).
52 Carmichael, S. T. & Price, J. L. Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal Comp Neurol* 363, 615-641, (1995).

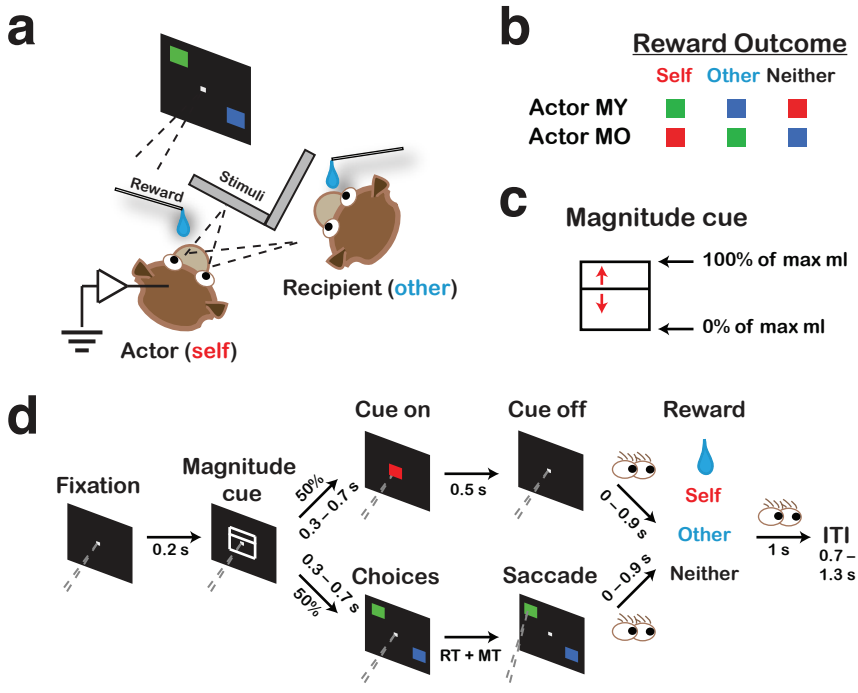

a
Reward
Stimuli
Recipient (other)
Actor (self)
b
Reward Outcome
Self
Other
Neither
Actor MY
Actor MO
c
Magnitude cue
100% of max ml
0% of max ml
d
Fixation
0.2 s
Magnitude cue
50%
0.3 – 0.7 s
Cue on
0.5 s
Cue off
0 – 0.9 s
Reward
Self
Other
Neither
1 s
ITI
0.7 – 1.3 s
0.3 – 0.7 s
50%
Choices
RT + MT
Saccade
0 – 0.9 s

Figure 1

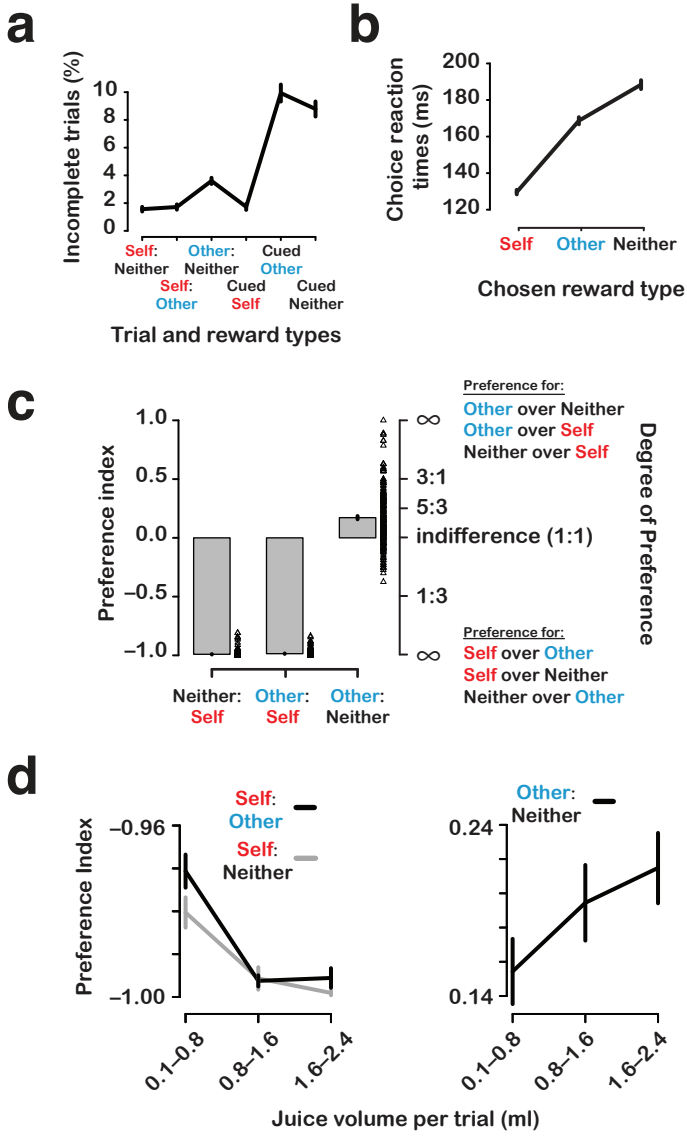

a
Incomplete trials (%)
10
8
6
4
2
0
Self: Neither
Other: Neither
Cued Other
Self: Other
Cued Self
Cued Neither
Trial and reward types
b
Choice reaction times (ms)
200
180
160
140
120
Self
Other
Neither
Chosen reward type
c
Preference index
1.0
0.5
0.0
−0.5
−1.0
Preference for:
Other over Neither
Other over Self
Neither over Self
∞
3:1
5:3
indifference (1:1)
1:3
∞
Preference for:
Self over Other
Self over Neither
Neither over Other
Degree of Preference
Neither: Self
Other: Self
Other: Neither
d
Preference Index
−0.96
−1.00
Self: Other
Self: Neither
0.24
0.14
Other: Neither
0.1–0.8
0.8–1.6
1.6–2.4
Juice volume per trial (ml)

Figure 2

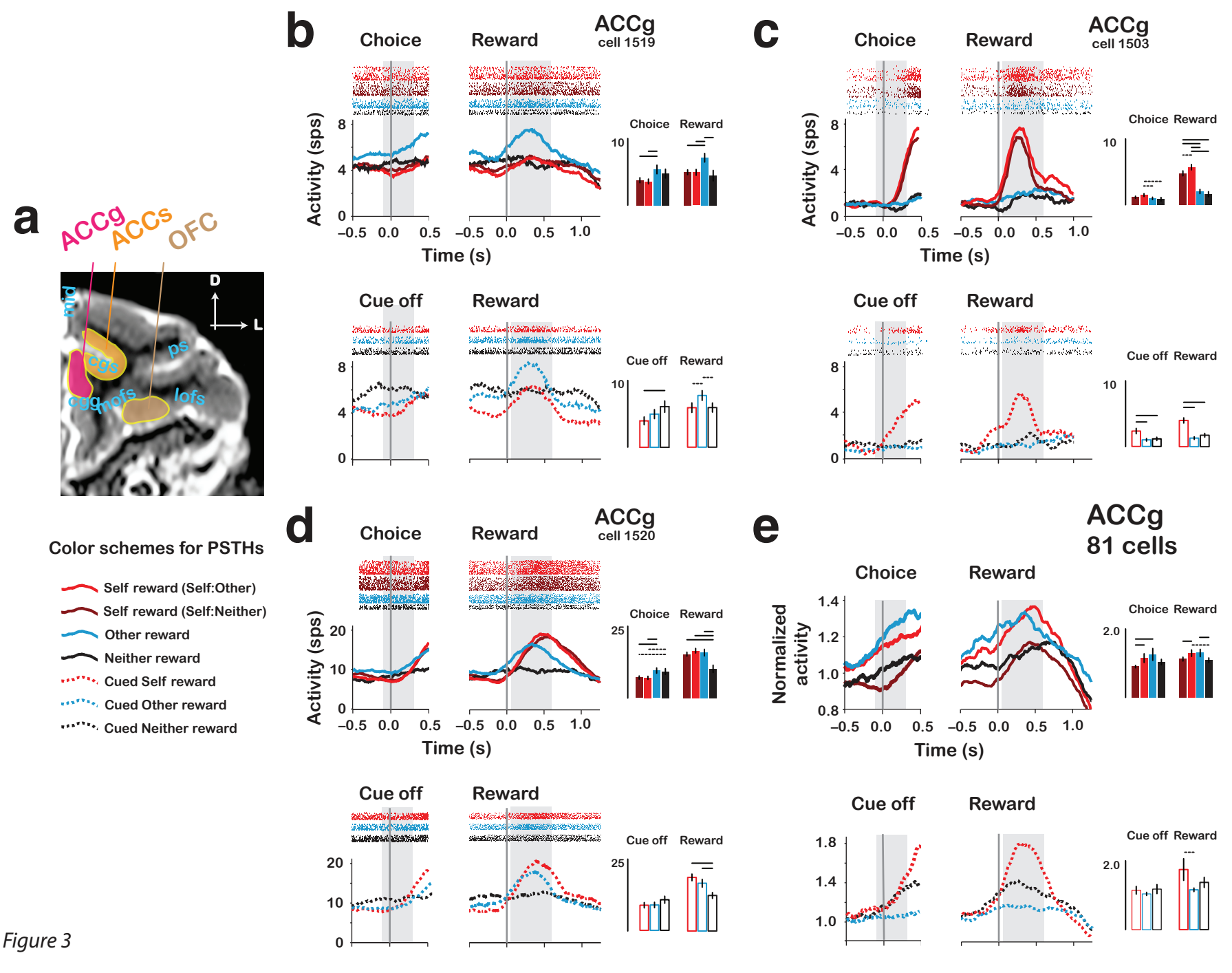

a
ACCg
ACCs
OFC
mfd
D
L
ps
cgs
cgg
ofs
lofs
Color schemes for PSTHs
Self reward (Self:Other)
Self reward (Self:Neither)
Other reward
Neither reward
Cued Self reward
Cued Other reward
Cued Neither reward
b
Choice
Reward
ACCg
cell 1519
Activity (sps)
Time (s)
Cue off
c
cell 1503
d
cell 1520
e
ACCg
81 cells
Normalized activity

Figure 3

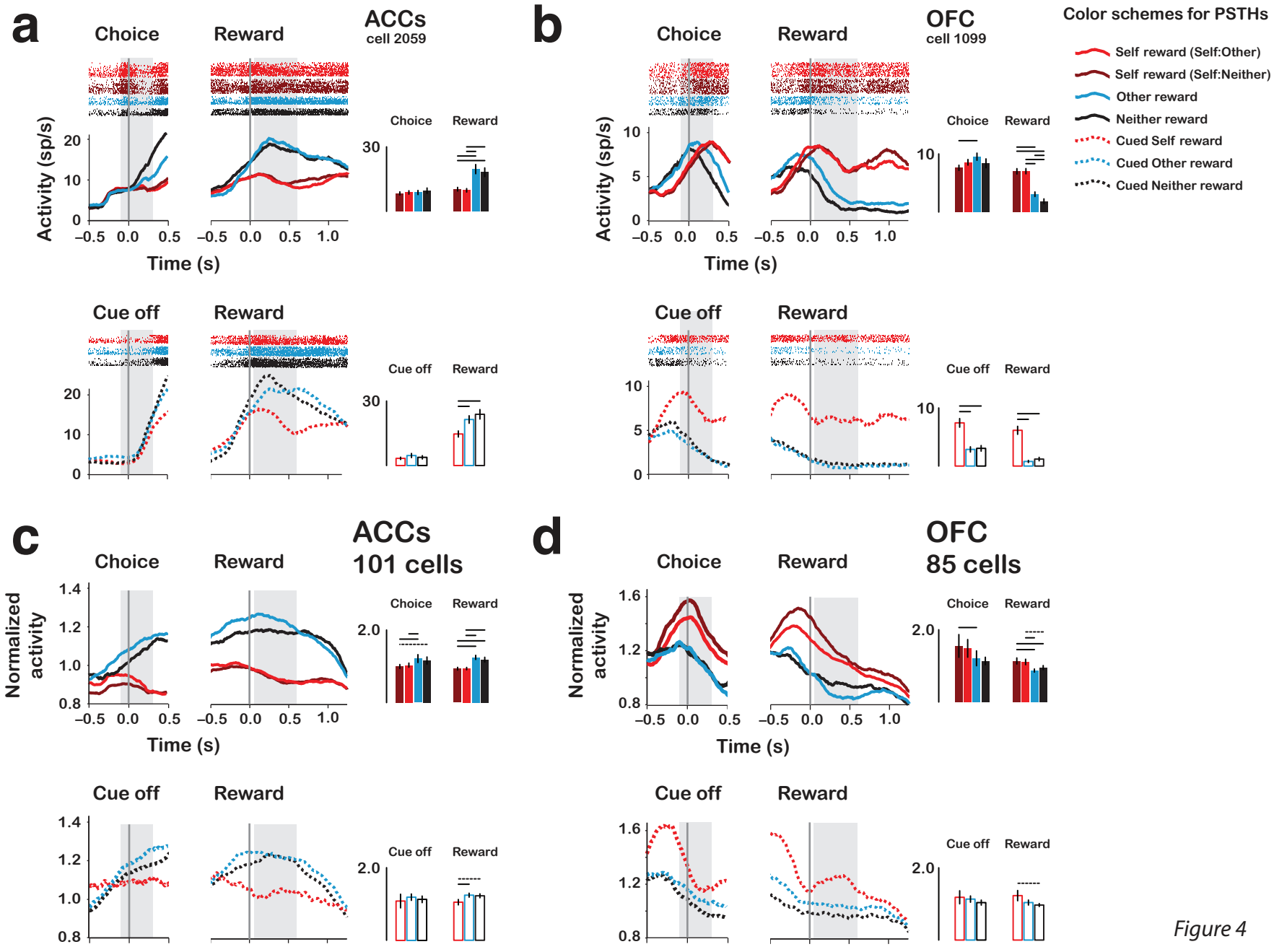

a
Choice
Reward
ACCs
cell 2059
Activity (sp/s)
Time (s)
Cue off
b
OFC
cell 1099
Color schemes for PSTHs
Self reward (Self:Other)
Self reward (Self:Neither)
Other reward
Neither reward
Cued Self reward
Cued Other reward
Cued Neither reward
c
ACCs
101 cells
Normalized activity
d
OFC
85 cells

Figure 4

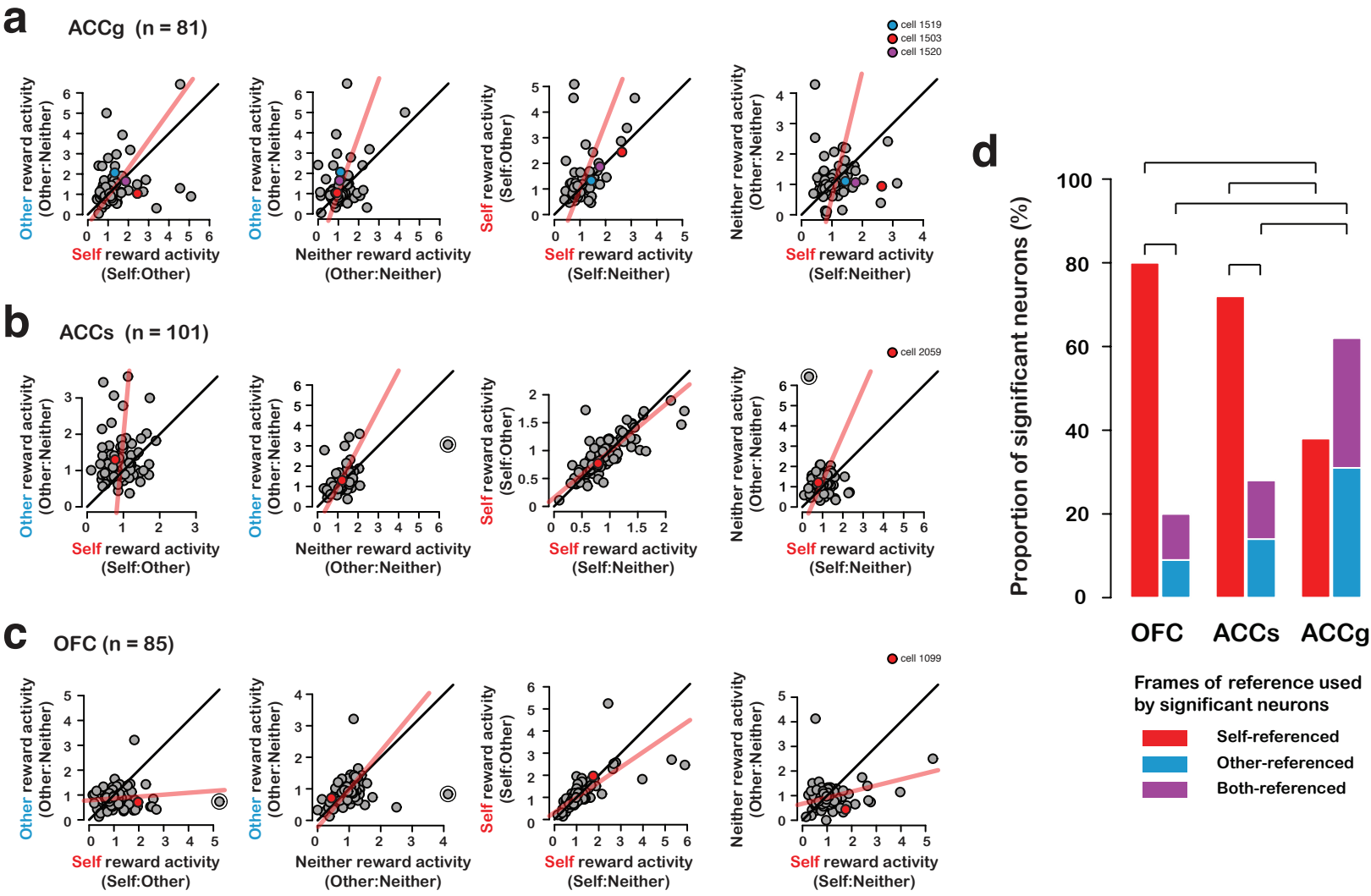

a
ACCg (n = 81)
cell 1519
cell 1503
cell 1520
Other reward activity (Other:Neither)
Self reward activity (Self:Other)
Other reward activity (Other:Neither)
Neither reward activity (Other:Neither)
Self reward activity (Self:Other)
Self reward activity (Self:Neither)
Neither reward activity (Other:Neither)
Self reward activity (Self:Neither)
b
ACCs (n = 101)
cell 2059
c
OFC (n = 85)
cell 1099
d
Proportion of significant neurons (%)
OFC
ACCs
ACCg
Frames of reference used by significant neurons
Self-referenced
Other-referenced
Both-referenced

Figure 5

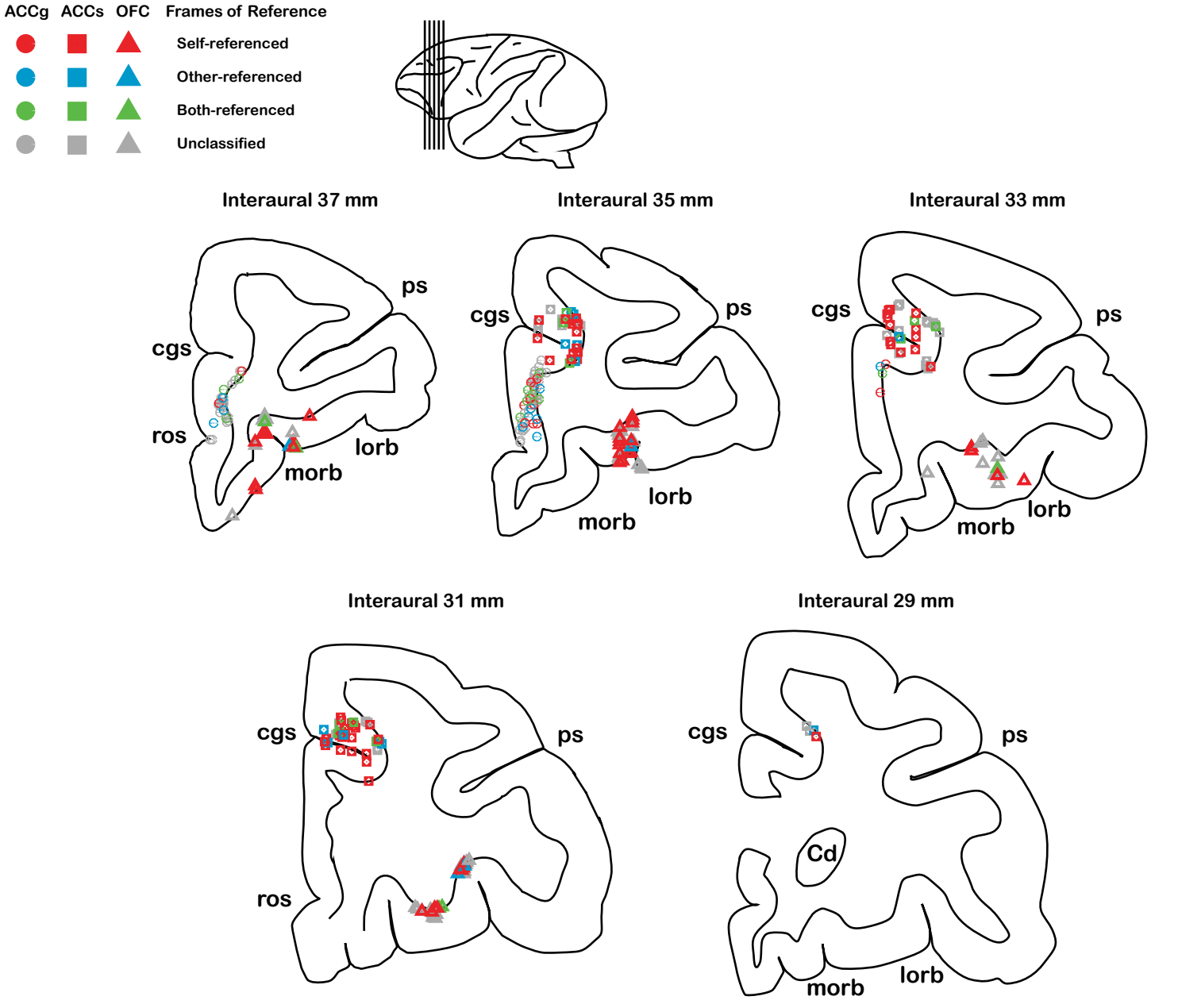

ACCg
ACCs
OFC
Frames of Reference
Self-referenced
Other-referenced
Both-referenced
Unclassified
Interaural 37 mm
ps
cgs
ros
lorb
morb
Interaural 35 mm
cgs
ps
lorb
morb
Interaural 33 mm
cgs
ps
morb
lorb
Interaural 31 mm
cgs
ps
ros
Interaural 29 mm
cgs
ps
Cd
morb
lorb

Figure 6

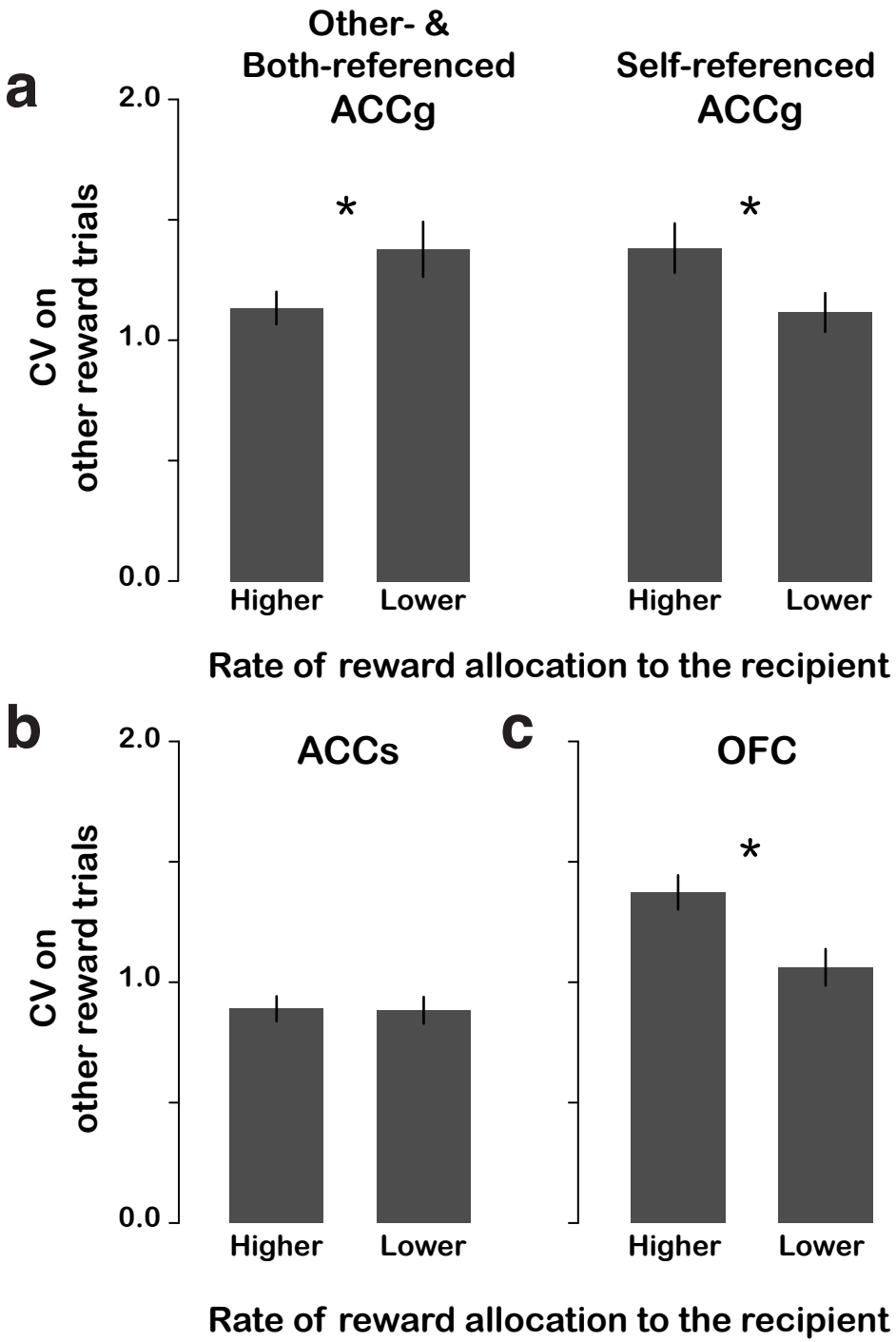

a
Other- &
Both-referenced
ACCg
Self-referenced
ACCg
2.0
1.0
0.0
CV on
other reward trials
*
*
Higher
Lower
Higher
Lower
Rate of reward allocation to the recipient
b
ACCs
c
OFC
2.0
1.0
0.0
CV on
other reward trials
*
Higher
Lower
Higher
Lower
Rate of reward allocation to the recipient

Figure 7

Supplementary Information

Neuronal reference frames for social decisions in primate frontal cortex

Steve W. C. Chang, Jean-Francois Gariépy, and Michael L. Platt

Supplementary Information Table: 1
Supplementary Information Figures: 9

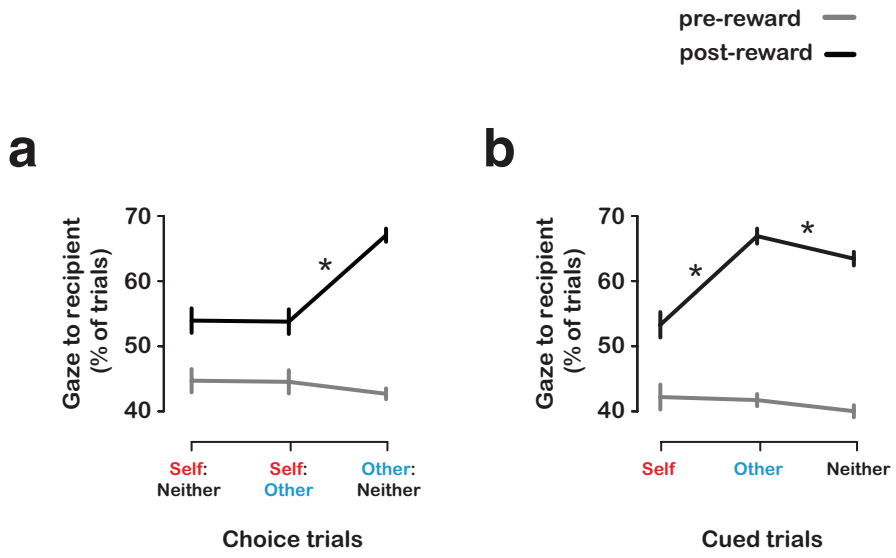
Supplementary Table 1

Supplementary Table 1. Classification of the reward type, trial type, and reward size selectivities at the level of individual neurons from OFC, ACCs, and ACCg, based on analysis of variance.

Area	Proportion of significant neurons by factors (reward epoch)		Proportion of significant neurons between different rewards (reward epoch)		Proportion of reference frame types (reward or choice/cue epoch)	
ACCg	*Reward Outcome*	57% (*n*=81)	*Self vs. Neither*	31% (*n*=81)	*SELF frame of reference*	15% (*n*=12/81) [**38%**] (*n*=12/32) (MY: 48%; MO: 18%)
	Trial Type	36% (*n*=81)				
	Reward Volume	12% (*n*=81)	*Self vs. Other*	36% (*n*=81)		
	Reward Outcome x Trial Type	30% (*n*=81)	*Other vs. Neither*	25% (*n*=81)	*OTHER frame of refernece*	12% (*n*=10/81) [**31%**] (*n*=10/32) (MY: 19%; MO: 55%)
	Reward Volume x Reward Outcome	7% (*n*=81)	*Self (Self:Other) vs. Self (Self:Neither)*	12% (*n*=81)		
	Reward Volume x Trial Type	4% (*n*=81)			*BOTH frame of reference*	12% (*n*=10/81) [**31%**] (*n*=10/32) (MY: 33%; MO: 27%)
	Reward Outcome x Trial Type x Reward Volume	7% (*n*=81)				
ACCs	*Reward Outcome*	72% (*n*=101)	*Self vs. Neither*	57% (*n*=101)	*SELF frame of reference*	51% (*n*=51/101) [**72%**] (*n*=51/71) (MY: 82%; MO: 61%)
	Trial Type	52% (*n*=101)				
	Reward Volume	25% (*n*=76)	*Self vs. Other*	53% (*n*=101)		
	Reward Outcome x Trial Type	36% (*n*=101)	*Other vs. Neither*	20% (*n*=101)	*OTHER frame of refernece*	10% (*n*=10/101) [**14%**] (*n*=10/71) (MY: 9%; MO: 20%)
	Reward Volume x Reward Outcome	16% (*n*=76)	*Self (Self:Other) vs. Self (Self:Neither)*	5% (*n*=101)		
	Reward Volume x Trial Type	4% (*n*=76)			*BOTH frame of reference*	10% (*n*=10/101) [**14%**] (*n*=10/71) (MY: 9%; MO: 19%)
	Reward Outcome x Trial Type x Reward Volume	8% (*n*=76)				
OFC	*Reward Outcome*	57% (*n*=85)	*Self vs. Neither*	37% (*n*=85)	*SELF frame of reference*	42% (*n*=36/85) [**80%**] (*n*=36/45) (MY: 86%; MO: 70%)
	Trial Type	45% (*n*=85)				
	Reward Volume	24% (*n*=62)	*Self vs. Other*	42% (*n*=85)		
	Reward Outcome x Trial Type	37% (*n*=85)	*Other vs. Neither*	14% (*n*=85)	*OTHER frame of refernece*	5% (*n*=4/85) [**9%**] (*n*=4/45) (MY: 7%; MO: 12%)
	Reward Volume x Reward Outcome	10% (*n*=62)	*Self (Self:Other) vs. Self (Self:Neither)*	13% (*n*=85)		
	Reward Volume x Trial Type	10% (*n*=62)			*BOTH frame of reference*	6% (*n*=5/85) [**11%**] (*n*=5/45) (MY: 7%; MO: 18%)
	Reward Outcome x Trial Type x Reward Volume	11% (*n*=62)				

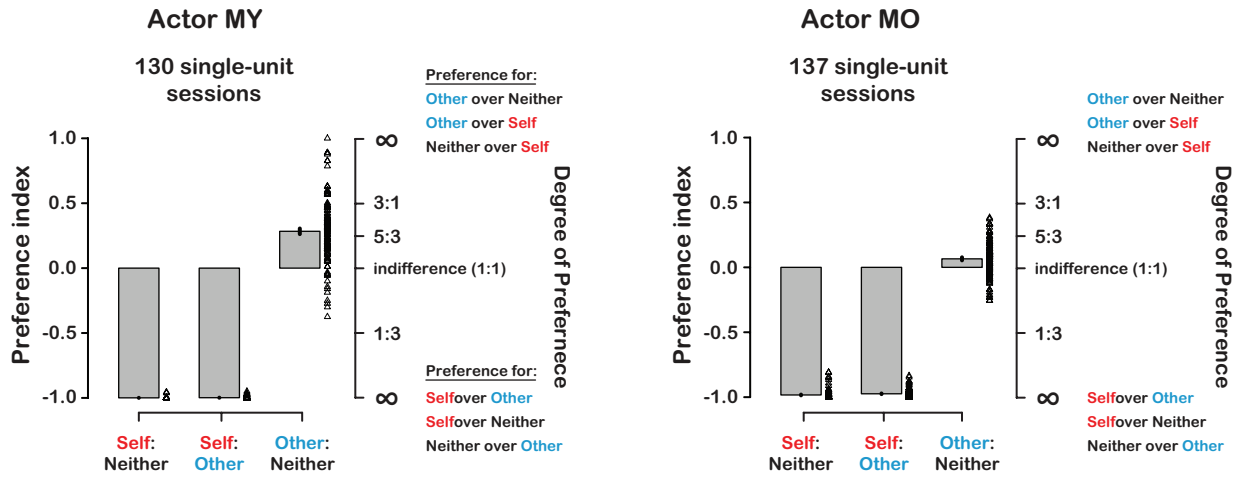
The bold percentages shown inside the brackets on the 4th column show the proportions out of *classfied* neurons. Shown below in parentheses are these proportions for each monkey. Significance in all panels was based on $P < 0.05$ (analysis of variance and tukey HSD tests).

Supplementary Figure & Legends



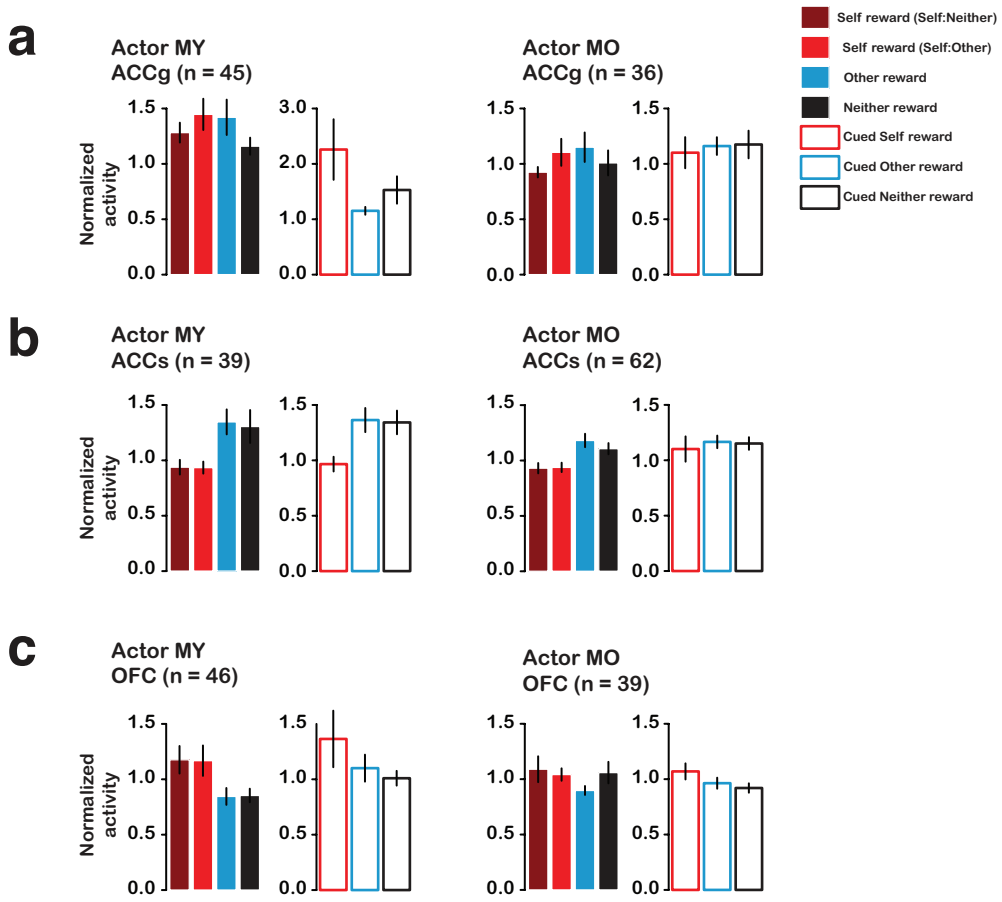
Supplementary Figure 1

Supplementary Figure 1 Percentage of gaze shifts (mean ± s.e.m.) directed toward the recipient prior to reward delivery (pre-reward epoch; grey) and following the onset of reward delivery (post-reward epoch; black) on choice trials (**a**) and cued trials (**b**). Asterisks indicate significant differences ($P < 0.05$, Welch two sample *t*-test). *Choice trials*. During the delay period between choice and reward delivery (pre-reward epoch, 0–0.9 sec; see **Fig. 1d**), gaze frequencies were comparable across trials involving actors' received and allocated rewards to other (44.75 ± 1.78% [mean ± s.e.m.], 44.57 ± 1.78%, 42.74 ± 0.83% on *Self:Neither*, *Self:Other*, *Other:Neither* trials, respectively; all comparisons $P > 0.17$, paired *t*-test). Following the onset of reward delivery (post-reward epoch, 1 sec), however, these frequencies were significantly higher on *Other:Neither* trials (67.01 ± 1.01%) compared to *Self:Other* (53.78 ± 1.88%) and *Self:Neither* trials (53.95 ± 1.88%) (both, $P < 0.0001$, paired *t*-test). *Cued trials*. During the delay period between cue offset and reward delivery (pre-reward epoch), gaze frequencies were comparable across cued *self*, cued *other* and cued *neither* trials (42.19 ± 1.90, 41.72 ± 0.92, and 40.03 ± 0.92, respectively; all comparisons $P > 0.20$). Following the onset of reward delivery (post-reward epoch), however, gaze frequencies were the highest for cued *other* trials (66.70 ± 1.15), compared to cued *self* (53.19 ± 1.93) or cued *neither* trials (63.26 ± 1.05) (both $P < 0.05$). Therefore, actors looked at the recipient at different rates depending on reward outcomes, as reported previously[33,34].



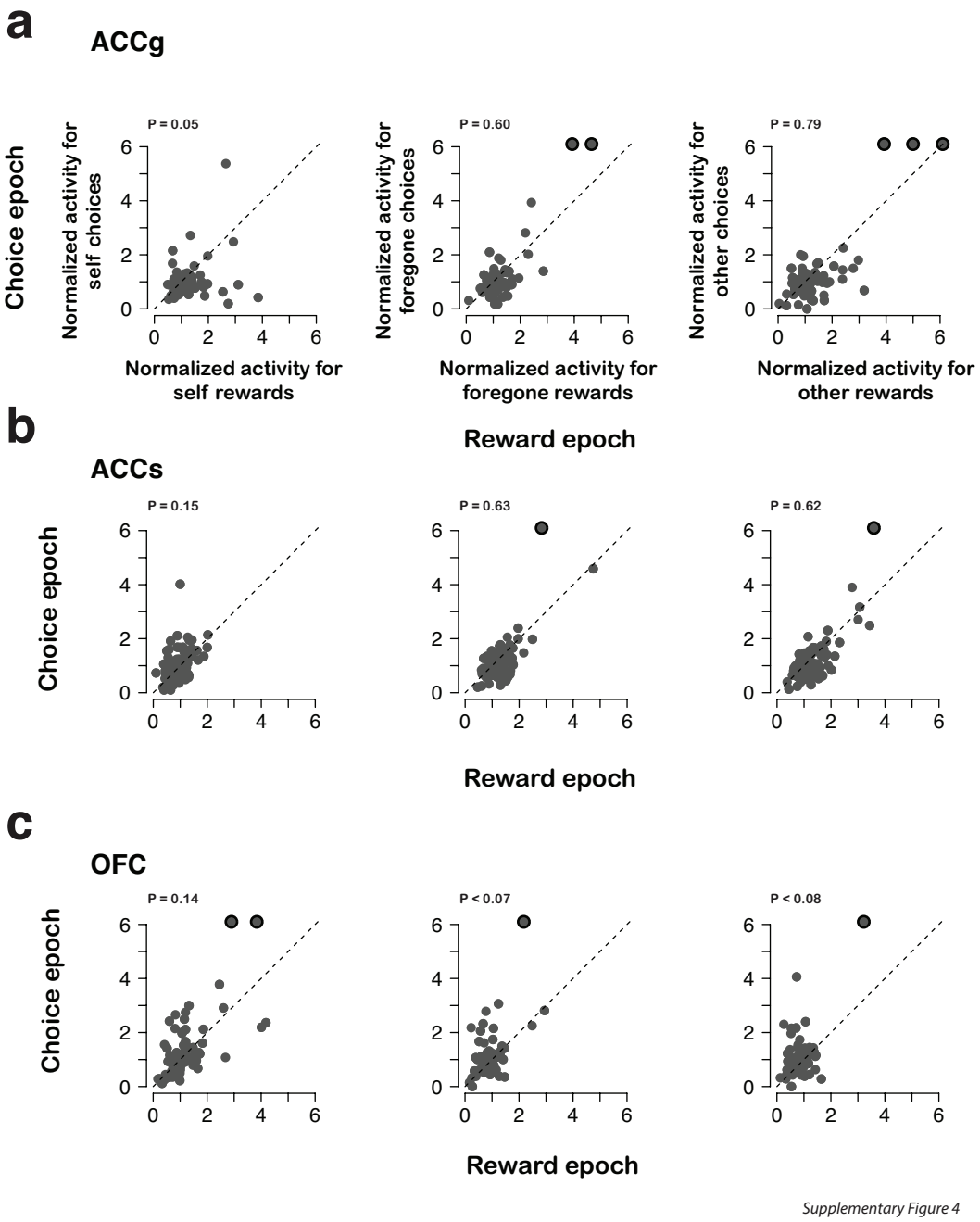
Supplementary Figure 2

Supplementary Figure 2 Choice preferences for each actor monkey (MY and MO). Shown are the preference indices as a function of reward outcome contrasts (i.e., choice contexts) for actor MY (left panel) (130 single-unit sessions) and for actor MO (right panel) (137 single-unit sessions). Data points next to each bar show individual sessions. The preference index for actor MY was –1.00 ± 0.00 (mean ± s.e.m.) for *Self:Neither*, –1.00 ± 0.00 for *Self:Other*, and 0.28 ± 0.02 for *Other:Neither* trials (significantly different from zero: all $P < 0.0001$, one-sample *t*-test). For actor MO, the preference index for *Self:Neither* was –0.98 ± 0.00, *Self:Other* was –0.97 ± 0.00, and *Other:Neither* was 0.07 ± 0.01 (significantly different from zero: all $P < 0.0001$, one-sample *t*-test). These choice behaviors are consistent with our previous studies using a similar behavioral paradigm, which also demonstrated differential reward allocation preferences depending on the familiarity and social status between the two animals[33] and the causal role of neuropeptide oxytocin in modulating these preferences[34].



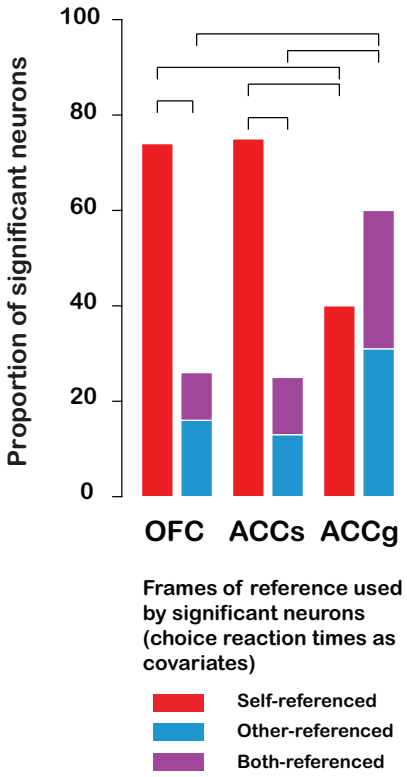
Supplementary Figure 3

Supplementary Figure 3 Preferential reward encoding biases in each actor monkey (MY and MO). Shown are the normalized responses (mean ± s.e.m.) to different reward outcomes during the reward epoch from ACCg (**a**), ACCs (**b**), and OFC (**c**). The inset shows the color coding scheme.



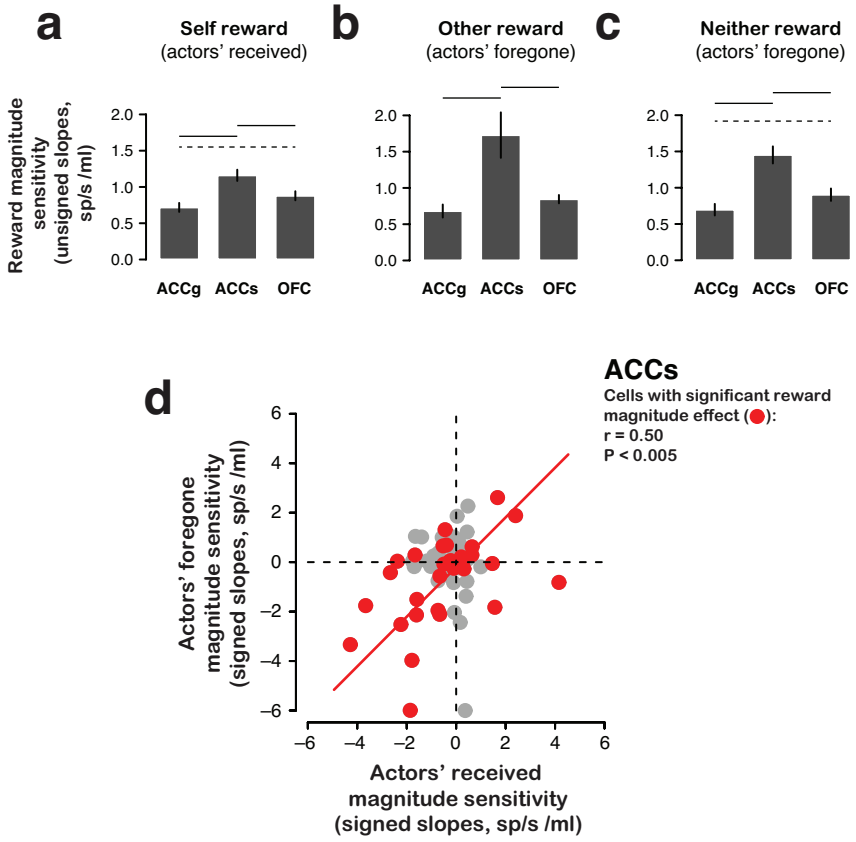
Supplementary Figure 4 Comparisons of neuronal responses across the choice/cue epoch and the reward epoch for (**a**) ACCg (n = 81), (**b**) ACCs (n = 101), and (**c**) OFC (n = 85). Plotted are the normalized responses from the two epochs for the following comparisons: *self* choices versus *self* rewards, foregone choices versus foregone rewards, and *other* choices versus *other* rewards. Data points with black outlines indicate that these values were truncated for the purpose of these displays. Significance values are shown at the top of each panel comparing the ordinate and abscissa at the population level using a paired t-test. At the population level, we observed a remarkable resemblance in neuronal activity across the choice/cue epoch and reward epoch in ACCg, ACCs, and OFC (**Fig. 3 & 4**). To quantify this similarity, we directly compared

normalized activity from the two epochs corresponding to the following pairs: *self* choices (choices leading to *self* rewards) versus *self* rewards (delivery of *self* rewards), foregone choices (leading to *other* or *neither* rewards) versus foregone rewards (delivery of *other* or *neither* rewards), and, finally, *other* choices versus *other* rewards. Although the majority of comparisons resulted in similar responses across the two epochs, there were some differences. The following summarizes the population level effects that we have observed here. ACCg as a population showed greater activity for *self* rewards during the time of reward delivery compared to the time around making choices leading to *self* rewards ($P = 0.05$, paired t-test). On the other hand, responses of ACCs neurons were similar in magnitude across all three comparisons (all $P > 0.15$, paired t-test). In contrast, OFC neurons showed a trend toward greater responses to foregone choices compared to foregone rewards ($P < 0.07$, paired t-test), as well as greater responses to *other* choices compared to *other* rewards ($P < 0.08$). At the individual cell level, however, a substantial number of neurons from each area showed significantly modulated activity across the two epochs. In ACCg, 38% ($n = 31$), 24% ($n = 19$), and 21% ($n = 17$) of neurons showed significantly modulated activity across the two epochs for *self*, *foregone*, and *other* choices versus rewards, respectively. In ACCs, these proportions were 63% ($n = 64$), 39% ($n = 39$), and 30% ($n = 30$), whereas in OFC, these proportions were 60% ($n = 51$), 40% ($n = 34$), and 28% ($n = 24$).



Supplementary Figure 5

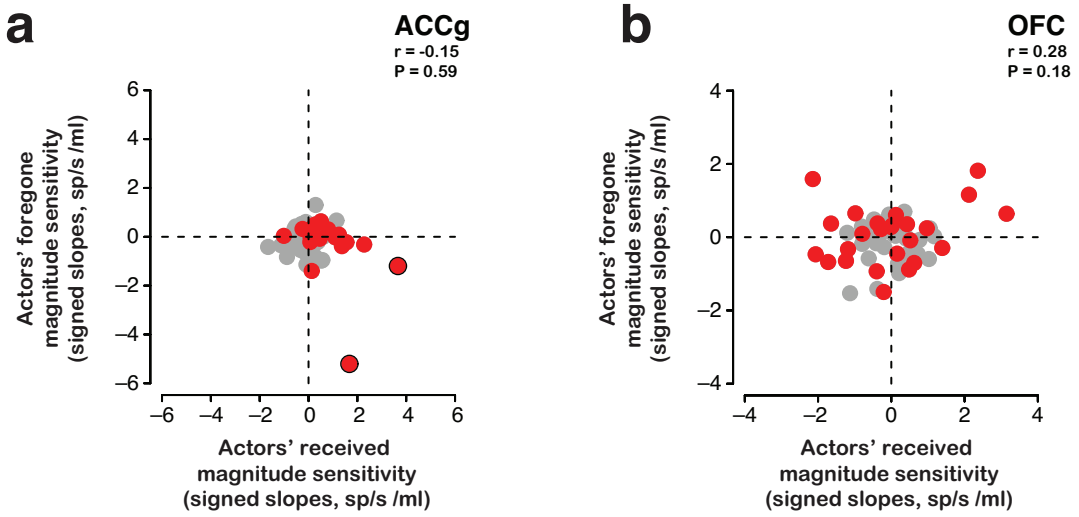
Supplementary Figure 5 Proportion of neurons (out of significantly classified neurons) after correcting for eye movement choice reaction times (i.e., by using reaction times as covariates in the general linear model) from OFC, ACCs, and ACCg using self-referenced, other-referenced, and both-referenced frames for representing reward outcomes. Inset shows color codes used in the bar graph. Bars indicate significant differences in proportions ($P < 0.05$, χ^2 test). The proportions of reference frame types across the three areas remain similar even after correcting for trial-by-trial reaction times (compare to **Fig. 5d**). **Figure 2b** clearly indicates that different choices are associated with different choice reaction times. Therefore, it is possible that differential encoding schemes reported here might be simply driven by the subjective value of different choices (as inferred from reaction times). For example, if neurons were merely computing the subjective value associated with different choices, one might expect choice reaction times to explain a large amount of variance in neuronal response. We directly tested this hypothesis by including trial-by-trial reaction times as covariates in the ANOVA and re-calculated the proportion of neurons classified within different functional categories. This figure shows the distribution of different reference frame types across each area after taking into account choice reaction times. The results are virtually identical to those shown in **Fig. 5d**, suggesting that self-referenced, other-referenced, and both-referenced neurons are not the products of encoding the subjective value (as revealed by reaction times) associated with different choices. Therefore, the neurons in the current study appear to signal specific decision outcomes during social decision-making, rather than directly encoding their subjective value.



Supplementary Figure 6

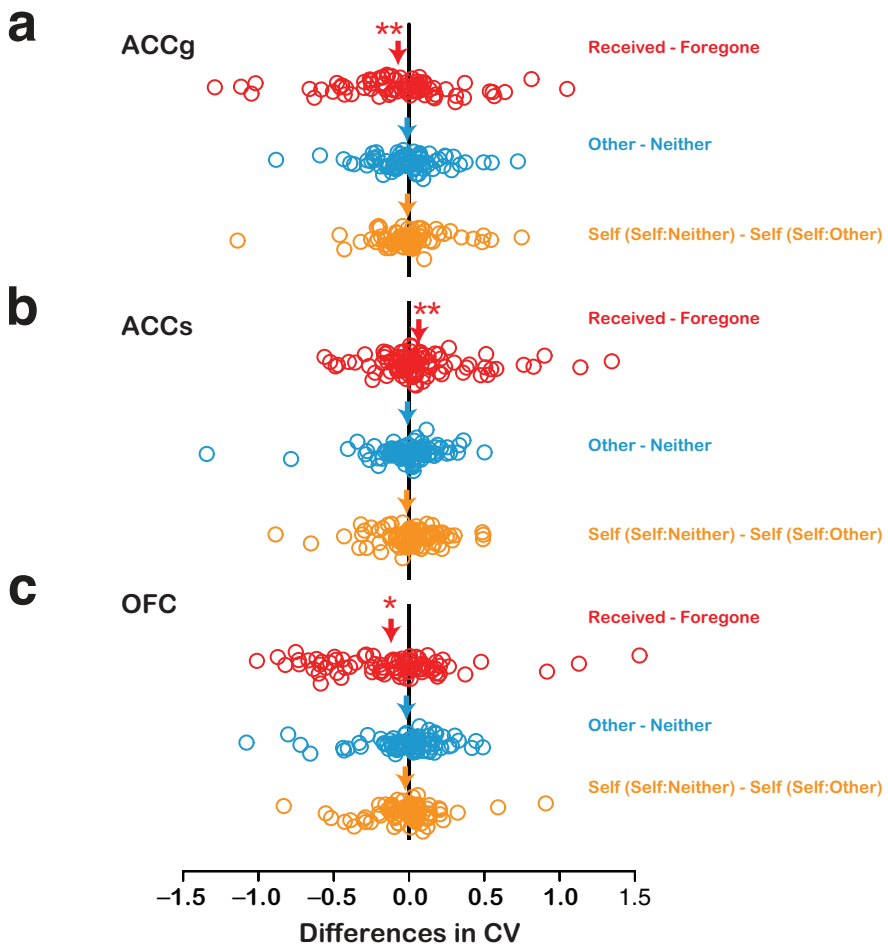
Supplementary Figure 6 Effects of value on *self*, *other*, and *neither* reward responses. (**a**–**c**) Reward magnitude sensitivity for ACCg, ACCs, and OFC (unsigned slopes, sp/s/ml, mean ± s.e.m.) computed from linear regression of reward epoch activity as a function of increasing value for *self* (**a**), *other* (**b**), or *neither* (**c**) outcomes. Horizontal bars above the histograms indicate significance (solid: $P < 0.05$, dashed: $P < 0.10$, Welch two sample *t*-test). (**d**) Value sensitivity for actors' received and foregone rewards are positively correlated in ACCs. Slopes of linear regressions of reward epoch activity as a function of reward volume for actors' received rewards (*Self:Neither*, *Self:Other*, and cued *self*) and foregone rewards (*other*, *neither* rewards, *other* cued, and *neither* cued). Red points indicate significant effects of reward value or interactions (ANOVA). Line indicates type II regression for neurons with significant reward value or interaction effects. *r* and *p* reflect Pearson's correlations for significant cells. Here we examined response modulations by the magnitude of reward delivered to *self*, *other*, and *neither*. Based on the ANOVA on reward epoch responses, 40% of OFC (out of 62 cells collected in a task with a reward magnitude cue), 40% of ACCs (of 76), and 21% of ACCg (of 81) showed

either a significant effect of reward magnitude or a significant interaction involving reward magnitude (**Supplementary Table 1**). ACCg contained a significantly smaller proportion of neurons modulated by reward magnitude compared to either OFC or ACCs (both $P < 0.05$, χ^2 test), whereas OFC and ACCs did not differ ($P = 1$). Out of these regions, ACCs neurons showed the greatest sensitivity to reward magnitude based on the slopes of the regression line for each neuron across all outcomes, consistent with a prominent role for ACCs in behavioral adjustment[35–37] in an environment with constantly changing reward types and contexts. We next explored in detail how the magnitude of foregone rewards and *self* rewards was encoded in each area. We found a significant positive relationship between actors' received and forgone rewards in the sample of ACCs neurons showing significant effects of reward magnitude (significant cells: $r = 0.50$, $P < 0.005$; all cells: $r = 0.33$, $P < 0.005$, Pearson's correlation) (**d**). By contrast, we did not observe this relationship in the ACCg or OFC neurons with significant reward magnitude effects: both regions $|r| < 0.28$, $P > 0.18$; all cells: both $r < 0.21$, $P > 0.11$; Pearson's correlation) (see **Supplementary Figure 7**). Thus, the ACCs, but not the ACCg or OFC, processes actors' direct and forgone rewards in a similar manner (i.e., scale in the same direction), consistent with its hypothesized role in learning from both experience and observation[8,11].



Supplementary Figure 7

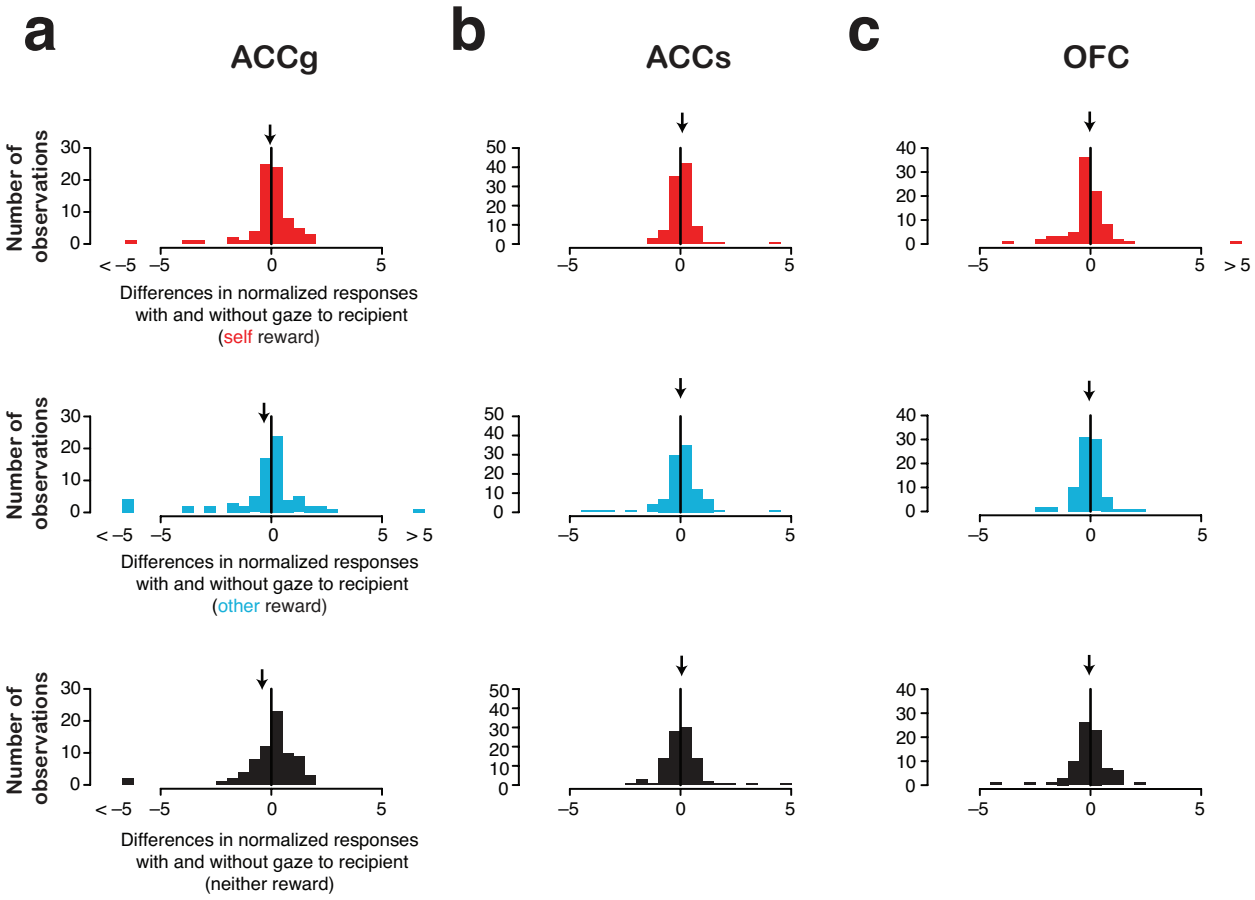
Supplementary Figure 7 Reward magnitude sensitivity between actors' received and foregone rewards in ACCg (**a**) and OFC (**b**). Plotted are slopes from a linear regression of reward epoch activity as a function of different reward volumes between actors' received rewards (*Self:Neither*, *Self:Other*, and cued *self*) and foregone rewards (*other*, *neither* rewards, *other* cued, and *neither* cued). Red data points indicate significant reward magnitude main or interactions effects (ANOVA). Shown as texts in the inset are *r* and significance from Pearson's correlation for the significant neurons. Data points with black outline show the outlier cells excluded from the correlation analysis.



Supplementary Figure 8

Supplementary Figure 8 Differences in coefficient of variation (CV) across different reward outcomes reflect reward bias in ACCg (**a**), ACCs (**b**) and OFC (**c**). Plotted are differences in CV between a pair of reward categories (as indicated on the right of each distribution). We compared individual neuron averages of all trials in which the actors received rewards against all trials in which the actors did not receive rewards (*Received – Foregone*) (top of each panel). We also compared individual neuron averages of trials in which the recipient received the rewards against trials in which no one received rewards (*Other – Neither*) (middle of each panel), and between trials in which the actors received rewards in *Self:Neither* against *Self:Other* contexts (*Self (Self:Neither) – Self (Self:Other)*) (bottom of each panel). If applicable, the data were collapsed across choice and cued trials for this analysis. Data points are jittered in the vertical dimensions for visibility. Asterisks above the data points indicate significance (**:$P < 0.05$, *: $P < 0.10$, one sample *t*-test) in the distribution. An alternative way to examine neuronal information encoding is to assess whether lower trial-to-trial variability is associated with preferred outcomes. We tested whether the coefficient of variation in firing rates (CV; **Online Methods** Eq. 2) was systematically lower for preferred reward outcomes (based on response magnitude) compared to

non-preferred reward outcomes. We found this to be the case. The OFC population showed a lower CV for *self* rewards vs. rewards delivered to *other* or *neither* (*Received – Foregone*, –0.12 ± 0.04 [mean ± s.e.m.], $P < 0.01$, one-sample *t*-test), whereas the ACCs population showed a lower CV for rewards delivered to *other* or *neither* (*Foregone*) (0.07 ± 0.03, $P < 0.05$, one sample *t*-test). In ACCg, where some neurons preferred *self* and some preferred *other* rewards, we found a lower CV only for actors' received rewards compared to foregone rewards (–0.07 ± 0.04, $P < 0.09$, one sample *t*-test, $P < 0.05$, bootstrap test), but no difference between *other* and *neither* rewards, or between the two contexts of receiving *self* rewards (all $P > 0.34$). Thus, the most robust responses of neurons in all three areas were also the most reliable.



Supplementary Figure 9

Supplementary Figure 9 Reward coding is not driven by gaze shifts directed at the recipient. Shown are histograms of the differences in normalized reward epoch responses between trials with gaze shifts and without gaze shifts (responses 'with' – responses 'without' gaze shifts), for trials in which rewards were delivered to *self* (top), *other* (middle), or *neither* (bottom), for ACCg (**a**), ACCs (**b**), and OFC (**c**) populations. Arrows indicate distribution means. In the reward-allocation task, actors were allowed to look at the recipient (**Fig. 1d** & **Supplementary Figure 1**). To rule out the possibility that preferential reward responses of neurons in these areas were simply driven by where the actors looked on a given trial, we compared reward epoch responses between trials *with* and *without* gaze shifts to the recipient. We found no systematic differences in these reward responses at the population level (each areas and each reward outcome: all $P > 0.20$, Wilcoxon signed rank test). The reward-related responses in the three regions are thus neither simply driven by preparation to look at the recipient nor elicited as a consequence of inspecting the recipient.

Behavioral and Brain Sciences

Games people (and monkeys) play: toward a neuroecology of social interaction.

--Manuscript Draft--

Manuscript Number:	
Full Title:	Games people (and monkeys) play: toward a neuroecology of social interaction.
Short Title:	Toward a neuroecology of social interaction
Article Type:	Commentary Article
Corresponding Author:	Michael L Platt Duke University UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Duke University
Corresponding Author's Secondary Institution:	
First Author:	Jean-Francois Gariépy
First Author Secondary Information:	
Order of Authors:	Jean-Francois Gariépy
	Steve W.C. Chang
	Michael L Platt
Order of Authors Secondary Information:	
Abstract:	In this issue, Schilbach et al. defend a "second-person neuroscience" perspective that focuses on the neural basis of social cognition during live, on-going interactions between individuals. We argue that a second-person neuroscience would benefit from formal approaches borrowed from economics and behavioural ecology and that it should be extended to social interactions in non-human animals.

Commentary

Target article: Schilbach, Timmermans, Reddy, Costall, Bente, Schlicht, Vogeley.

Word counts: Abstract (55), Main text (870), References (811), Entire text (1842).

Commentary title: Games people (and monkeys) play: toward a neuroecology of social interaction

Jean-François Gariépy[a,b], Steve W. C. Chang[a,b], Michael L. Platt[a,b,c,d]

[a]Department of Neurobiology, Duke University School of Medicine, Durham, NC 27701
[b]Center for Cognitive Neuroscience, Duke University, Durham, NC 27708
[c]Department of Evolutionary Anthropology, Duke University, Durham, NC 27708
[d]Department of Psychology and Neuroscience, Duke University, Durham, NC 27708

Mailing address: Michael Platt, B243F Levine Science Research Center, Duke Box 90999
Phone: 919-668-0332
Emails: jeanfrancois.gariepy@gmail.com, steve.chang@duke.edu, platt@neuro.duke.edu
Homepage: http://www.neuro.duke.edu/faculty/platt/

Abstract

In this issue, Schilbach et al. defend a "second-person neuroscience" perspective that focuses on the neural basis of social cognition during live, on-going interactions between individuals. We argue that a second-person neuroscience would benefit from formal approaches borrowed from economics and behavioural ecology and that it should be extended to social interactions in non-human animals.

Main text

The "second-person neuroscience" proposed by Schilbach et al. in the current issue of Behavioral and Brain Sciences proffers the intriguing idea that social cognition during real-time interactions with another individual may be fundamentally different from passive observations of another's actions. Understanding the contribution of neural processes to on-going interactions with complex beings is a fascinating research direction, with potential implications for the treatment of disorders attended by social deficits as well as for ethics and public policy.

Several decades of neuroscientific research have sketched out the neural circuits that may translate perceptual information about other individuals into purposeful action. Specifically, regions of the human and nonhuman primate brain including the superior temporal sulcus and fusiform face area contribute to social identification (Tsao et al., 2008). The ventromedial prefrontal cortex, orbitofrontal cortex and striatum appear to play a role in translating knowledge of others into motivational signals (Burke et al., 2010; Cooper et al., 2010; Azzi et al., 2012). The anterior cingulate cortex and fronto-insular cortex contribute to empathy and other-regarding cognition (Decety, 2010; Gu et al., 2010). The so-called mentalizing and mirroring networks appear to participate in action and intention understanding (Rizzolatti and Sinigaglia, 2010; Becchio et al., 2012). Circuits connecting these areas could translate

social perceptual information into appropriate actions via decision-making mechanisms (Baumgartner et al., 2009; Knoch et al., 2009; Baumgartner et al., 2011).

To better understand the neural mechanisms underlying social cognition, we propose that social neuroscience needs to ground its predictions and hypotheses in a formal framework such as that provided by behavioural game theory (Platt and Glimcher, 1999; Dorris and Glimcher, 2004; Kosfeld et al., 2005; Tomlin et al., 2006; Lee, 2008; Gintis, 2009). Schilbach et al. criticize game theoretical approaches for not recreating the dynamics of everyday real-life social encounters, but this common opposition has been rebutted before (Gintis, 2009). Game theoretical frameworks are general and open, allowing formal delineation of specific hypotheses while not imposing restrictions on the behaviours that are being described. Formal approaches borrowed from economics, game theory and behavioural ecology have been extremely useful in describing decisions in dynamic foraging or social environments (Sugrue et al., 2004; Lee, 2008; Hayden et al., 2011; Chang et al., 2011).

These approaches can be extended to describe the dynamics of interacting individuals, with several advantages. First, they allow us to generate empirically-testable and mathematically-formalizable predictions about the neural mechanisms that could underlie decisions in complex social environments. Second, they allow for comparative analyses of decision processes in humans and other animals with respect to the demands placed on them in specific physical and social environments (Kacelnik and Bateson, 1996; Stephens et al., 2002; Heilbronner et al., 2008).

Schilbach et al. also raise the concern that classical game theory paradigms involve mainly one-shot interactions or turn-taking. Although this structure is often used for simplicity, we contend that continuous interactions in interactive games can also be effectively described using a similar theoretical framework (Debreu, 1952; Braun et al., 2009). Such mathematical tools would help translate some of the intuitive aspects of Schilbach et al.'s approach into concrete experimental predictions.

Second-person neuroscience would also benefit from broadening its inquiry to the interactions of nonhuman animals (Washburn et al., 1990; Fujii et al., 2007; Chang et al., 2011). Social complexity appears to have favored the evolution of higher social cognition in animals that have brains similar to ours, like macaques (Barsalou, 2005; Rudebeck et al., 2006; Tsao et al., 2008; Azzi et al., 2012; Chang et al., 2012) and in animals that have very different brains as well, like scrub jays and rooks (Emery and Clayton, 2001; Bird and Emery, 2010). We know from research in macaques, sheep and mice that social cognition in mammals appears to rely on neural circuits that are similar, and perhaps homologous, to those in humans (Barsalou, 2005; Rudebeck et al., 2006; Tsao et al., 2008; Sanchez-Andrade and Kendrick, 2009; Jeon et al., 2010; Azzi et al., 2012). One possible explanation is that we inherited those circuits from a common ancestor that possessed some level of social complexity. Alternatively, similar constraints applying to neural circuits could also have caused them to evolve in similar ways to support similar functions. How such functions are accomplished by neural circuits in animals with brains that are very different from our own — such as birds — remains an open question.

We agree with Schilbach et al. that studying the neural processes mediating live interaction between real agents is crucial for the maturation of social neuroscience as a discipline. What we propose is to

supplement this approach with formal game theory and value-based analysis of preferences in humans and nonhuman animals. In our lab, for example, we study pairs of monkeys interacting both in economical and interactive games (for instance, see Chang et al., 2011). Estimating preferences allows us to quantify how much monkeys value certain options (for instance, giving juice to another monkey). Game theory will allow us to generate predictions of the equilibriums that could develop over time between two interacting individuals (see Braun et al., 2009). Understanding the neural processes that underlie social cognition in such animals could powerfully inform our understanding of the evolutionary origins of our own social abilities.

References

Azzi, J.C., Sirigu, A. & Duhamel, J.R. (2012) Modulation of value representation by social context in the primate orbitofrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America* 109:2126-2131.

Barsalou, L.W. (2005) Continuity of the conceptual system across species. *Trends in Cognitive Sciences* 9:309-311.

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K. & Fehr, E. (2009) The neural circuitry of a broken promise. *Neuron* 64:756-770.

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C. & Fehr, E. (2011) Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience* 14:1468-1474.

Becchio, C., Cavallo, A., Begliomini, C., Sartori, L., Feltrin, G. & Castiello, U. (2012) Social grasping: From mirroring to mentalizing. *Neuroimage* 61:240-248.

Bird, C.D. & Emery, N.J. (2010) Rooks perceive support relations similar to six-month-old babies. *Proceedings in Biological Sciences of the Royal Society* 277:147-151.

Braun, D.A., Ortega, P.A. & Wolpert, D.M. (2009) Nash equilibria in multi-agent motor interactions. *PLoS Computational Biology* 5:e1000468.

Burke, C.J., Tobler, P.N., Schultz, W. & Baddeley, M. (2010) Striatal BOLD response reflects the impact of herd information on financial decisions. *Frontiers in Human Neuroscience* 4:48.

Chang, S.W., Barter, J.W., Ebitz, R.B., Watson, K.K. & Platt, M.L. (2012) Inhaled oxytocin amplifies both vicarious reinforcement and self reinforcement in rhesus macaques (Macaca mulatta). *Proceedings of the National Academy of Sciences of the United States of America* 109:959-964.

Chang, S.W., Winecoff, A.A. & Platt, M.L. (2011) Vicarious reinforcement in rhesus macaques (macaca mulatta). *Frontiers in Neuroscience* 5:27.

Cooper, J.C., Kreps, T.A., Wiebe, T., Pirkl, T. & Knutson, B. (2010) When giving is good: ventromedial prefrontal cortex activation for others' intentions. *Neuron* 67:511-521.

Debreu, G. (1952) A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences of the United States of America* 38: 886–893.

Decety, J. (2010) The neurodevelopment of empathy in humans. *Developmental Neuroscience* 32:257-267.

Dorris, M.C. & Glimcher, P.W. (2004) Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* 44:365-378.

Emery, N.J. & Clayton, N.S. (2001) Effects of experience and social context on prospective caching strategies by scrub jays. *Nature* 414:443-446.

Fujii, N., Hihara, S. & Iriki, A. (2007) Dynamic social adaptation of motion-related neurons in primate parietal cortex. *PLoS One* 2:e397.

Gintis, H. (2009) *The bounds of reason: Game theory and the unification of the behavioral sciences.* Princeton University Press. New Jersey, USA.

Gu, X., Liu, X., Guise, K.G., Naidich, T.P., Hof, P.R. & Fan, J. (2010) Functional dissociation of the frontoinsular and anterior cingulate cortices in empathy for pain. *Journal of Neuroscience* 30:3739-3744.

Hayden, B.Y., Pearson, J.M. & Platt, M.L. (2011) Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience* 14:933-939.

Heilbronner, S.R., Rosati, A.G., Stevens, J.R., Hare, B. & Hauser, M.D. (2008) A fruit in the hand or two in the bush? Divergent risk preferences in chimpanzees and bonobos. *Biology Letters* 4:246-249.

Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H.E., Lin, S.Y., Rabah, D., Kinet, J.P. & Shin, H.S. (2010) Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. *Nature Neuroscience* 13:482-488.

Kacelnik, A. & Bateson, M. (1996) Risky theories: The effects of variance on foraging decisions. *American Zoology* 36:402-434.

Knoch, D., Schneider, F., Schunk, D., Hohmann, M. & Fehr, E. (2009) Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proceedings of the National Academy of Sciences of the United States of America* 106:20895-20899.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U. & Fehr, E. (2005) Oxytocin increases trust in humans. *Nature* 435:673-676.

Lee, D. (2008) Game theory and neural basis of social decision making. *Nature Neuroscience* 11:404-409.

Platt, M.L. & Glimcher, P.W. (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400:233-238.

Rizzolatti, G. & Sinigaglia, C. (2010) The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience* 11:264-274.

Rudebeck, P.H., Buckley, M.J., Walton, M.E. & Rushworth, M.F. (2006) A role for the macaque anterior cingulate gyrus in social valuation. *Science* 313:1310-1312.

Sanchez-Andrade, G. & Kendrick, K.M. (2009) The main olfactory system and social learning in mammals. *Behavioural Brain Research* 200:323-335.

Stephens, D.W., McLinn, C.M., Stevens & J.R. (2002) Discounting and reciprocity in an Iterated Prisoner's Dilemma. *Science* 298:2216-2218.

Sugrue, L.P., Corrado, G.S. & Newsome, W.T. (2004) Matching behavior and the representation of value in the parietal cortex. *Science* 304:1782-1787.

Tomlin, D., Kayali, M.A., King-Casas, B., Anen, C., Camerer, C.F., Quartz, S.R. & Montague, P.R. (2006) Agent-specific responses in cingulate cortex during economic exchanges. *Science* 312:1047-1050.

Tsao, D.Y., Moeller, S. & Freiwald, W.A. (2008) Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America* 105:19514-19519.

Washburn, D.A., Hopkins, W.D. & Rumbaugh, D.M. (1990) Effects of competition on video-task performance in monkeys (Macaca mulatta). *Journal of Comparative Psychology* 104:115:121.

Wilson, E.O. (1975) *Sociobiology: The New Synthesis.* Harvard University Press. Cambridge, Massachusetts, USA.